Final report on the evaluation of the Growth Model Pilot Project

U.S. DEPARTMENT OF EDUCATION

Final Report on the Evaluation of the Growth Model Pilot Project

Final Report on the Evaluation of the Growth Model Pilot Project

---

Prepared by:

Thomas B. Hoffer
E. C. Hedberg
Kevin L. Brown
Marie L. Halverson
Paki Reid-Brossard

National Opinion Research Center (NORC)
at the University of Chicago

Andrew D. Ho, Harvard Graduate School of Education

Katherine Furgol, University of Iowa

Prepared for:

U.S. Department of Education
Office of Planning, Evaluation and Policy Development
Policy and Program Studies Service

2011

This report was prepared for the U.S. Department of Education under Contract Number ED-06-CO-0053 with National Opinion Research Center at the University of Chicago. Andrew Abrams served as the contracting officer's representative. The views expressed herein do not necessarily represent the positions or policies of the Department of Education. No official endorsement by the U.S. Department of Education is intended or should be inferred.

U.S. Department of Education
Arne Duncan
Secretary

Office of Planning, Evaluation and Policy Development
Carmel Martin
Assistant Secretary

Policy and Program Studies Service
Adriana de Kanter
Acting Director

January 2011

To obtain copies of this report:

Write to ED Pubs, Education Publications Center, U.S. Department of Education, NTIS, P.O. Box 22207, Alexandria, VA 22304

Or **fax** your request to 703-605-6794.

Or **e-mail** your request to edpubs@edpubs.ed.gov.

Or **call** in your request toll free to 1-877-433-7827 (1-877-4-ED-PUBS). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 1-877-576-7734. If 877 service is not available in your area, call 1-800-872-5327 (1-800-USA-LEARN).

Or **order online** at http://edpubs.ed.gov.

This publication is also available at the Department's website at http://www.ed.gov/about/offices/list/opepd/ppss/reports.html#title.

On request, this publication is available in alternate formats, such as Braille, large print or computer diskette. For more information, contact the Department's Alternate Format Center at 202-260-0852 or 202-260-0818.

Contents

Exhibits

Acknowledgments

This report benefited from the efforts and thoughtful input of many individuals. Andrew Abrams at the U.S. Department of Education's Policy and Program Studies Service (PPSS) directed the project and provided helpful guidance and encouragement at each stage. Daphne Kaplan, David Goodwin, and Stephanie Stullich, also at PPSS, offered much constructive advice and feedback throughout the project.

The data analyzed in this report were drawn from the U.S. Department of Education's ED*Facts* archive and from each of the separate state departments of education participating in the Growth Model Pilot Project. We are grateful to Ross Santy and Susan Thompson-Hoffman for their generous help and patience with the ED*Facts* data. The Alaska, Arizona, Arkansas, Delaware, Florida, Iowa, North Carolina, Ohio, and Tennessee state departments of education were very responsive to our various requests for data and information about their pilot programs, and we would particularly like to acknowledge the help of Les Morse, Robert Franciosi, Denise Airola, Robin Taylor, Ed Croft, Thomas Deeter, Kenneth Barbour, Matt Cohen, and Dan Long.

We were very fortunate to receive thoughtful and timely reviews from several individuals both within and outside the U.S. Department of Education. In particular, we received extremely helpful guidance in the early stages of the project and detailed feedback on early drafts of the report from Patrick Rooney at the Office of Elementary and Secondary Education (OESE) within ED. We also thank Abigail Potts and Victoria Hammer of OESE for sharing their expertise in the design and review stages. In addition we would like to thank Alan Ginsburg, Jay Noell, and Emma Vadehra for their insight during initial drafting of the final report. Vadehra and Scott Sargrad provided very helpful comments on the later stages of the final report.

The project had a Technical Working Group (TWG) consisting of David Figilo, Thomas Fisher, Pete Goldschmidt, Rachel Quenemoen, and Keith Zvoch. The TWG provided guidance with the study design and analysis plans, and contributed many useful comments and corrections on the Interim Report. Pete Goldschmidt reviewed an early draft of this Final Report and provided many helpful suggestions.

The Final Report differed from the Interim Report in incorporating data from the 2007–08 school year and conducting an in-depth analysis of differences among the pilot states, particularly the effects of using different types of growth models. The additional analysis benefitted greatly from the expertise of Andrew Ho of the Harvard Graduate School of Education and Katherine Furgol of the University of Iowa, who drafted much of Chapter 4 and provided excellent suggestions for improving much of the rest of the report. Jennifer Dunn at Measured Progress also consulted on the design of the Final Report analyses. At NORC, the project was helped at several key junctures by the insightful perspectives of Stephen Raudenbush and Larry Hedges. Marie Halverson and Sarah Kay McDonald directed the project. Cindy Simko contributed greatly to collecting data from the states and organizing and implementing the many reviewer comments. The task of formatting this report in each of its iterations was handled with great skill and efficiency by Imelda Demus, Isabel Guzman-Barron, and Yajaira Gijon.

Executive Summary

The U.S. Department of Education (ED) initiated the Growth Model Pilot Project (GMPP) in November 2005 with the goal of approving up to ten states to incorporate growth models in school adequate yearly progress (AYP) determinations under the *Elementary and Secondary Education Act* (*ESEA*). After extensive reviews, nine states were fully approved for the initial phase of the pilot project by the 2007–08 school year: Alaska, Arizona, Arkansas, Delaware, Florida, Iowa, North Carolina, Ohio, and Tennessee. Based on analyses of data provided by the U.S. Department of Education and by the pilot grantee states, this report describes the progress these states made in implementing the GMPP in the 2007–08 school year.

GMPP Objectives

Use of growth models for determining AYP is attractive to states and local districts because it offers a means to identify schools in which students are making progress even though they may not yet be reaching proficiency standards. Without recognition of the progress made by these students, these schools would be subject to school improvement actions that may not be appropriate in light of their demonstrable improvements.

The standard method of determining AYP has been the "status model," in which school performance is mainly evaluated in terms of the proportion of students meeting or exceeding proficiency standards for reading and mathematics. The status model has been supplemented with "safe-harbor" provisions. Sometimes referred to as an improvement model, safe-harbor recognizes schools that do not make AYP under the status model as making AYP if the percentage of non-proficient students decreased by 10 percent or more from the previous to the current school year.

In contrast, growth models measure how much students have gained from one year to the next using longitudinal records of individual student achievement in reading and mathematics. The models determine whether each student is "on-track" to reach or exceed the state's grade-level proficiency cut points (or thresholds) on the annual tests of reading and mathematics within three or four years or by a specified grade level (usually grade eight or nine) as defined by the state's particular growth model. For purposes of determining AYP, a student who is not proficient but on-track can be counted the same as a proficient student or as some fraction thereof.

Consistent with the general rules of *ESEA* accountability, the GMPP requires that the data on students' proficiency and on-track to proficiency results are used to assess all students and each reporting subgroup: major racial or ethnic groups (American Indians, Asians, blacks, Hispanics, whites), students from low-income households, students with disabilities, and students with limited English proficiency. Each group must meet the same annual measurable objectives (AMOs) in order for a school to make AYP.

States Included in the Pilot Project

The GMPP began in 2005 with two states—North Carolina and Tennessee—approved to use growth models for *ESEA* accountability in the 2005–06 school year. The initial pilot project was limited to include no more than 10 states. The number approved to implement growth models under the pilot expanded to eight states in 2006–07 and nine states in 2007–08: Alaska, Arizona, Arkansas, Delaware, Florida, Iowa, North Carolina, Ohio, and Tennessee. This report focuses only on the nine states approved under the initial pilot project for 2007–08.[1]

In December 2007, the U.S. Department of Education removed the initial pilot program's cap of 10 states and all states were able to use growth models for AYP determinations, pending approval by the Department. Following the change, six more states were approved under the open application process: Michigan and Missouri (beginning in the 2007–08 school year); and Colorado, Minnesota, Pennsylvania, and Texas (beginning in the 2008–09 school year). These six states were not part of the initial pilot project and were thus beyond the scope of this report.

Findings

Features of Growth Models Implemented by Pilot States

The growth models implemented under the GMPP were all designed to augment rather than replace the standard status model and safe-harbor provisions for determining school AYP. The growth models resulted in more schools making AYP than would have been the case using only status and safe-harbor.

Eight of the nine states used growth criteria only after schools failed to make AYP under the status and safe-harbor provisions. Delaware was the exception, applying growth results before status and safe-harbor. The designs of the pilots in the other eight states applied growth criteria only to *ESEA* reporting groups that did not reach their annual measurable objectives (AMOs) or obtain AYP via safe-harbor provisions. Furthermore, in all but Florida and Tennessee, the growth criteria were applied *within* the *ESEA* reporting groups only to the students who did not reach the proficiency threshold. The number of non-proficient but on-track students was added to the number of proficient students and the reporting group was counted as meeting the AMO if the total was high enough.

[1] An evaluation by the Office of Elementary and Secondary Education of growth model results in North Carolina and Tennessee for the 2005–06 school year is available at http://www.ed.gov/admins/lead/account/growthmodel/gmeval0109.doc.

Comparing AYP by Growth with Status and Safe-Harbor

The growth models in the nine states that provided data resulted in some schools making AYP that would not have made AYP under status or safe-harbor alone (see Exhibit S.1).[2] Among all schools, 9 percent made AYP in 2007–08 uniquely because of the growth model (that is, they did not make AYP by status or safe-harbor). This compares with 3 percent of all schools making AYP due to growth during the 2006–07 school year. However, most of the schools that made AYP by growth were located in Ohio; excluding Ohio, only 2 percent of all schools in the other eight states made AYP by growth. The percentages of all schools that made AYP uniquely by growth varied widely among the states, ranging from literally 0 or 1 percent of all schools in Alaska, Arizona, and North Carolina to 34 percent of all schools in Ohio. The impact of GMPP varied little across the two school years with the exception of Iowa, which saw a decline from 11 percent to just 2 percent of schools making AYP because of the pilot program.

Ohio entered the program in 2007–08 and saw its growth model account for 50 percent of the schools that made AYP. Much of the high rate of making AYP by growth observed in Ohio is likely explained by their procedures for identifying non-proficient students as being on-track to proficiency. To reduce the risk of misclassifying those students as being not on-track, Ohio adopted a much more inclusive definition of on-track to proficiency and this had the consequence of helping relatively large numbers of schools make AYP that would not have done so without the growth model.

[2] Delaware did not report AYP results in this form to the U.S. Department of Education because it elected to apply growth model results before status or safe-harbor. However, Delaware provided the authors with additional data identifying which schools uniquely made AYP because of the availability of the growth model and they are used in Exhibits S.1 and S.2.

Exhibit ES.1
Percentage of Schools That Made AYP Before and After the Application of the Growth Model, in Nine States, in 2007–08 and 2006–07

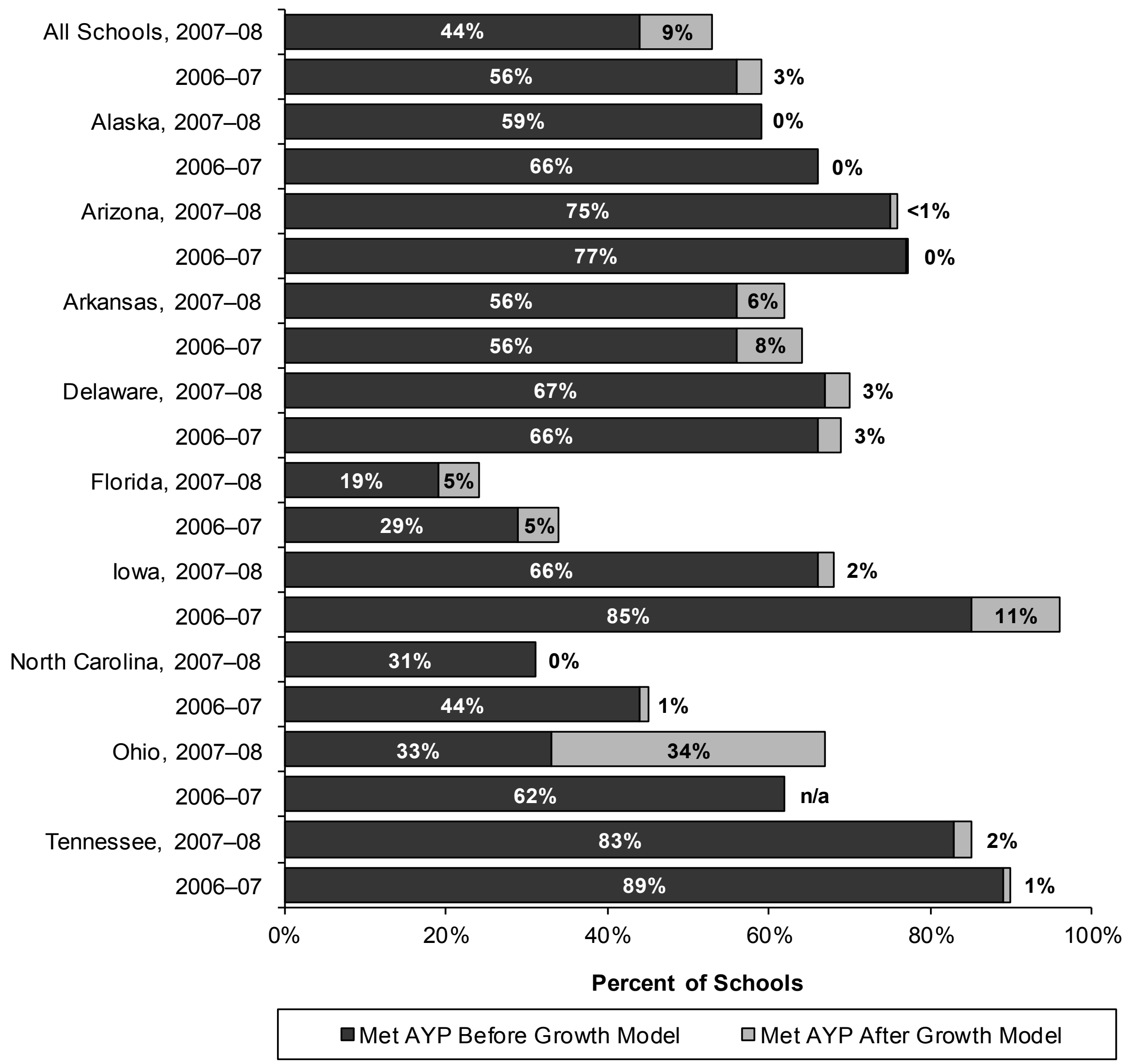


Exhibit reads: Among all schools containing grade levels for which growth models were used for AYP determinations in the nine pilot states during the 2007–08 school year, 44 percent made AYP using status or safe-harbor methods. Another 9 percent made AYP because of the growth model.

As the bar chart illustrates, the pilot states also varied greatly in the proportion of their schools that made AYP under status or safe-harbor. The 1,246 schools making AYP uniquely by growth represented a percentage increase in the schools making AYP (column E in Exhibit S.2) of 20 percent across all states, but this increase was largely comprised of 983 schools in Ohio and 153 schools in Florida. Excluding Ohio, the overall percentage increase due to the GMPP in the other eight states was 5 percent.

Another measure of GMPP impact is the extent to which it reduced the number of schools that would not have made AYP had the growth model not been available. From this perspective, the number of schools that did not make AYP by either status or safe-harbor was reduced by 16 percent overall because of the GMPP (column F in Exhibit S.2). The percentage reduction was by far the highest in Ohio (50 percent), followed by Arkansas (13 percent), and Tennessee (10 percent). Excluding Ohio, the overall percentage reduction due to the GMPP in the other eight states was 4 percent.

Exhibit ES.2
Number of Schools Making or Not Making AYP By Status, Safe-Harbor or Growth Model, Percentage Increase in Number of Schools That Made AYP Due to Growth, and Percentage Decrease in Number of Schools That Did Not Make AYP Due to Growth, by State, 2006–07

Pilot States	A: Number of Schools Making AYP by Status or Safe-Harbor	B: Number of Schools Making AYP by Growth	C: Number of Schools not Making AYP	D: Total Number of Schools (A+B+C)	E: Percentage Increase in Schools Making AYP Due to Growth (B/A)	F: Percentage Decrease in Non-AYP Schools Due to the Growth Model (B/(B+C))
All Nine States	6,213	1,246	6,617	14,076	20%	16%
Alaska	292	0	203	495	0%	0%
Arizona	1,117	8	371	1,496	1%	2%
Arkansas	500	52	338	890	10%	13%
Delaware	123	5	55	183	4%	8%
Florida	632	153	2,495	3,280	24%	6%
Iowa	721	23	354	1,098	3%	6%
North Carolina	737	0	1,612	2,349	0%	0%
Ohio	961	983	984	2,928	102%	50%
Tennessee	1,130	22	205	1,357	2%	10%

Exhibit reads: The 1,246 schools that made AYP by growth increased the number of schools making AYP from 6,213 to 7,459 schools, which was a percentage increase of 20 percent. Of the schools that did not make AYP under either status or safe-harbor (1,246+6,617=7,863), the growth model decreased the non-AYP total by 16 percent.

Source: U.S. Department of Education, ED*Facts* and the Delaware state department of education.

Impact of Growth Models on AYP Rate Among High-Poverty Schools

Schools serving disadvantaged populations have been found to make AYP at much lower rates than those serving more affluent populations (U.S. Department of Education, 2007). The growth model pilots may reduce these associations to some extent by identifying high-growth schools serving low-income and minority communities.

The results of this analysis showed that schools serving economically disadvantaged student populations in all pilot states except for Arkansas were more likely than more-advantaged schools to make AYP by growth. Across all nine states, the percentage increase in the number of high-poverty schools making AYP as a result of the growth model being available was 20 percent (994 schools instead of 826 schools), compared to 18 percent among low-poverty schools (2,298 schools instead of 1,946 schools). However, the percentage increases among high-poverty schools were much greater than those among low-poverty schools in Florida (77 instead of 45 for high-poverty schools compared to 300 instead of 272 for low-poverty schools), Ohio (172 instead of 64 for high-poverty schools compared to 771 instead of 458 for low-poverty schools), and Tennessee (280 instead of 264 for high-poverty schools compared to 120 instead of 119 for low-poverty schools).

Hypothetical Results of Using Growth Models Instead of Status and Safe-Harbor

The GMPP application guidelines noted that states could use growth as the primary accountability indicator for all schools, including those that made AYP by status. This growth-only model would lower AYP percentages by excluding proficient students who are not on-track to maintain proficiency due to declines. A possible advantage of using a growth-only model is that it could provide better predictions of whether students will reach or maintain proficiency goals within the pilot's time frame (three or four years, or by grade 8 or 9 in most of the states). The student data were used to assess the extent to which the schools that made AYP by status and safe-harbor (examined separately) would also have met or exceeded their AMO for reading and mathematics proficiency if the growth criteria of on-track-to-proficiency were used instead of the status or safe-harbor criteria.

Overall, 62 percent of the schools that made AYP by status criteria in the nine states also would have met their reading and mathematics AMOs strictly by using the growth criteria. Results varied widely among the states, ranging from only 46 percent in Arizona and 47 percent in Arkansas and North Carolina to 75 percent or more in Ohio, Delaware, and Tennessee. These findings provide some evidence that despite the low rates of making AYP by growth found in most states, many schools had sufficiently high rates of students being on-track to reach or maintain proficiency to meet their AMOs using growth only. Under the normal practice of applying status before growth, the large numbers of schools that could have met their AMOs by growth-only were obscured. The percentage of schools that made AYP by safe-harbor and that also met or exceeded their reading and mathematics AMOs under the growth-only criteria was much lower (28 percent overall). Across the eight states with safe-harbor schools (Delaware had none), the percentages did not exceed 30 percent in any state except Arkansas (64 percent) and Ohio (45 percent). This relatively low level of overlap of safe-harbor and growth-only outcomes indicates that, despite an ostensibly similar purpose of identifying progress toward proficiency goals among schools that have not reached status model AMOs, the two methods often led to different results, possibly because safe-harbor measures school progress while growth-only measures individual student growth.

Effects of Growth Model Types on Student and School Outcomes

The growth models implemented by the nine pilot states generally grouped into three types: *transition matrix* models (which evaluate student progress from year to year in terms of a relatively small set of discrete performance levels), *trajectory* models (which use the gap between a baseline test score and a performance standard several years out to calculate the amount of growth required to become proficient), and *projection* models (which use current and past test scores to statistically predict performance several years ahead). Effects of model types were assessed by using data from a single state and applying generic versions of the models to those data.

Results show that the projection model functions in stark contrast with transition matrix and trajectory models in terms of identifying students as on-track to reach or maintain proficiency. The transition matrix model acts as a coarse, categorical approximation of the trajectory model, with agreement rates on classifications of students as on-track or not on-track to proficiency of over 90 percent. The overlap is greatest when time horizons for the trajectory model are long or when the number of categories for the transition matrix model is large. In contrast, when a projection model classifies a non-proficient student as on-track, the probability that a transition matrix or trajectory model agrees is near zero, and vice versa. However, agreement rates for non-proficient students are around 60 percent due to the three types of models agreeing about non-proficient students who are not on-track.

The underlying reasons for the contrast between the projection model and the transition matrix or trajectory models relate to differing objectives guiding the models. Projection models are designed to maximize the accuracy of a prediction about whether a student will meet or exceed future proficiency standards. From a statistical standpoint, the best predictor of future achievement is past achievement. As a result, relatively few students with records of low achievement but evidence of improvement are predicted to meet or exceed future proficiency standards, while students with records of high achievement but evidence of slipping are very likely to be predicted to meet or exceed future proficiency standards. Transition matrix and trajectory models, in contrast, are designed to identify specific growth targets that each student must attain in order to meet or exceed the future proficiency standards, given his or her benchmark score (usually from the last test administration). Reflecting these different guiding objectives, projection models classify fewer previously non-proficient students who are gaining as on-track than the other models but a larger number of previously proficient decliners than the other models. For status-plus-growth models (which use growth model results only for non-proficient students), projection models will have the least impact, affecting only 10 to 20 percent of eligible (non-proficient) students while transition matrix and trajectory models affect over 20 percent. States with higher proficiency cut scores will have heightened differences between the model types and a lower proportional impact of projection models on eligible students.

In simulations of school AYP determinations, the models do not yield large differences in the percentages of schools making AYP when non-proficient students who are on-track to proficiency according to each type of model are added to the numbers of students meeting or exceeding the proficiency cut point (i.e., status-plus-growth). The models differ much more when a growth-only calculation is made first, followed by status-plus-growth calculations. Under the growth-only simulation using a realistic proficiency cut score and standard AMOs, very few schools would make AYP with a trajectory or transition matrix model while most schools would make AYP with a projection model.

Predictive Accuracy of the Different Types of Growth Models

The growth models implemented in the GMPP have two important goals that turn out to be somewhat contradictory: to predict as accurately as possible whether non-proficient students will attain proficiency within a delimited time frame, and to identify as clearly as possible the performance levels students must achieve at each grade level in order to attain proficiency within the designated time frame. A comparison of the student on-track versus not-on-track determinations from the three generic types of growth models using the same data shows that the projection model has the highest correct classification rates for future proficiency: over 80 percent. These rates are 5 to 20 percentage points higher than trajectory and transition matrix models, depending on the grade level and proximity to the growth model time limit.

While the greater predictive accuracy of the projection model may be cited as an advantage, the simpler trajectory and transition matrix models may provide clearer guidance to schools, teachers, students, and their parents about the amount of growth needed to reach or maintain proficiency from year-to-year. This is because the simpler models identify a level of achievement that each student must attain at each grade in order to be on-track to reach or maintain proficiency, while the projection model cannot easily identify intermediate achievement targets for individuals because of the complex statistical apparatus used to predict future scores.

Effects of Alternative Standards of Adequate Yearly Growth

The GMPP core principles dictated that the pilot growth models must be designed to assess progress toward grade-level proficiency. However, attaining the on-track designation can require very large annual growth increments for students who start at low levels of achievement. At the other extreme, for students who start at high levels, attaining the on-track designation may be possible with no growth at all. In light of these shortcomings of proficiency-based growth standards and current policy interest in alternatives, an analysis of the effects of other ways of assessing growth was undertaken.

An alternative criterion-referenced standard of annual growth that can be used with vertical test score scales is the difference between the proficiency cut scores in successive grade levels. Students gaining that amount or more would be considered to make "adequate yearly growth" regardless of whether they are proficient or on-track to become proficient. A simulation shows that the overall percentages of students meeting that alternative standard of growth are lower than the percentage of proficient students in both reading and mathematics, but that the percentages of non-proficient students meeting the alternative standard are higher than that meeting the GMPP standard. Consequently, adding the non-proficient students who met the alternative growth standard to the pool of proficient students would increase the overall rates of students who are arguably performing adequately (i.e., proficient or making reasonable progress over the past year). As expected, these methods are dependent on the location of proficiency cut scores up the vertical scale.

Qualifications and Implications

The results of this analysis show that use of growth models generally added to the number of schools making AYP but that the numbers were not large in almost all pilot states.

The GMPP growth models were required to measure students' growth toward meeting or exceeding proficiency standards in reading and mathematics. This means that substantial individual student performance improvements that do not reach the growth models' proficiency targets (generally within three or four years, or by grades 8 or 9) or subgroup- and school-aggregate student performance improvements that do not reach the proficiency targets (AMOs) are not recognized by the GMPP growth models. However, if *ESEA* regulations were revised so that a broader range of student gains were recognized as acceptable, more students and schools would be identified as making adequate progress than is currently the case. Such other targets generally involve use of a more finely graduated set of performance outcomes than proficiency or on-track to proficiency.

The generally low rates of making AYP by growth also reflect the impact of the various other (nongrowth) methods for determining AYP available in those states for schools to make AYP by status and safe-harbor (e.g., confidence intervals and multiyear averaging), such that the status and safe-harbor methods picked up schools which would have made AYP by growth had those various provisions not been available. Those additional nongrowth ways of making AYP are generally intended to reduce the chance of misclassifying schools as not making AYP but may obscure the extent (or lack) of student progress in the schools.

Within the current regulatory context of the *ESEA*, an implication of the results presented here is that states could clarify each school's progress by applying growth criteria to all their schools and subgroups before status and safe-harbor. The main advantage of applying the growth model before the status and safe-harbor models is that it would identify schools that are realizing adequate progress toward universal proficiency. This would clearly distinguish those schools from schools making AYP under status or safe-harbor criteria but not realizing growth sufficient to continue meeting their AMOs. Identifying such schools would serve as an early warning mechanism of possible problems. The exploratory analyses in this report also indicate that applying growth criteria before safe-harbor could usefully reclassify many (28 percent overall) of the current safe-harbor schools as making AYP by growth, and would clearly identify those that are not on-track to proficiency and thus likely headed for improvement status in the near future.

Another reporting option is to classify each school in terms of both growth and status. Schools making AYP would be distinguished as making AYP by both growth and status, by growth only, by status only, by a mix of status and growth, or by safe-harbor only. This would have the advantage of uniquely identifying different sets of schools (those making AYP in terms of both growth and status).

This study has also shown that the types of growth models states select for federal accountability purposes are consequential and raise some potentially difficult theoretical questions for policymakers. Projection models are likely to be more accurate than transition matrix or trajectory models in terms of predicting students' future attainment of proficiency targets, but the simpler trajectory and transition matrix models may provide clearer guidance on annual achievement goals to schools, teachers, students, and their parents.

The projection model is explicitly designed to provide probabilistic predictions whereas the other models do not. As a probabilistic estimator, the projection model carries a measure of uncertainty for each student's predicted score. An important issue illustrated by the Ohio model is whether to adjust a student's predicted score for the uncertainty and, if so, to what extent. The adjustment used by Ohio (adding two standard error units to each student's predicted score) was selected in order to make it highly unlikely that a student who actually was on-track was misclassified as not on-track. However, the adjustment also had the effect of classifying a much higher percentage of non-proficient students as on-track than the other pilot states. Careful consideration of the trade-off between false-negatives and false-positives is needed when adjustments for uncertainty are made to statistically derived growth model results.

I. Introduction

A key goal of the 1994 reauthorization of the *Elementary and Secondary Education Act (ESEA)* was to introduce a standards-based accountability system that "required states to define criteria for measuring *adequate yearly progress* (AYP) in school performance for Title I schools and districts."[3] States were given considerable latitude in how to determine AYP, with the majority relying on various types of aggregate school-improvement models rather than setting absolute proficiency targets for students.[4] However, the 2001 reauthorization of *ESEA* established the primacy of absolute proficiency standards by requiring states to: (a) develop grade-level specific proficiency standards in both reading and mathematics for grades 3–8 and one or more high school grades, and (b) to assess the performance of all students in those grades each year. The states' targeted percentages of students scoring at or above the proficiency standards increase at least every three years with all students expected to be proficient by 2014. Based on its students' scores, every public school is evaluated to determine whether or not it is making AYP, and consequences are applied to schools not making AYP for more than two consecutive years.

With the increasing availability of statewide longitudinally linked student performance records since 2001, it became possible for some states to measure student growth and use those data for accountability purposes. The U.S. Department of Education initiated the Growth Model Pilot Project (GMPP) in November 2005 with the goal of approving up to ten states to incorporate growth models in school AYP determinations under *ESEA*. Growth models are defined as complements or alternatives to the standard status model for determining school AYP. The status model bases AYP on the proportion of a school's students attaining proficiency in reading and mathematics in a given year. Growth models, in contrast, base AYP in part on the proportion of individual students who are making sufficient annual progress to reach grade-level proficiency within a specific time horizon of three to five years or by grades 7, 8, or 9. Growth models promise to provide a fuller understanding of school effectiveness and the progress each school's students are making toward their proficiency goals. The main objectives of the GMPP are to help states develop and implement models for determining school-level AYP that incorporate measures of student growth.

Scope of This Report

This final evaluation of the GMPP is restricted to the nine states approved for participation in the initial pilot project during the 2007–08 school year. The states approved for participation in the initial pilot project were: North Carolina and Tennessee (approved for implementation

[3] U.S. Department of Education, Office of the Under Secretary, Policy and Program Studies Service (2004), "Evaluation of Title I Accountability Systems and School Improvement Efforts (TASSIE): First-Year Findings," Washington, D.C. (retrieved December 2009 from http://www.ed.gov/rschstat/eval/disadv/tassie1/index.html).

[4] Ibid. Examples of school-improvement models under the 1994 reauthorization included making AYP if any growth in school-average achievement, or the percentage of proficient students, occurred from year-to-year or if the gap between low- and high-achieving students was reduced by a given percentage each year. However, a few states adopted absolute standards of proficiency and measured school progress in ways very similar to those mandated by the 2001 reauthorization of *ESEA*.

beginning in the 2005–06 school year); Alaska, Arizona, Arkansas, Delaware, Florida, and Iowa (approved for implementation beginning in the 2006–07 school year); and Ohio (approved to begin implementation in the 2007–08 school year). While the GMPP began in 2005 with a goal of approving up to 10 states, the Department made the option of using growth models available to all states in December 2007, substantially expanding the scope of the pilot. In June 2008 and January 2009 Secretary Spellings announced the approval of six additional growth models, in Michigan and Missouri (approved for use in the 2007–08 school year), and Colorado, Minnesota, Pennsylvania, and Texas (beginning in the 2008–09 school year), but those states were not part of the initial pilot project and thus are beyond the scope of this report.[5]

The report is designed to answer three questions:

How have states in the pilot project implemented growth models?

How does each pilot state's growth model affect the number and kinds of schools that make AYP?

What are the implications of the pilot project experience for extending and strengthening growth models within the context of *ESEA*?

The remainder of this chapter describes how the GMPP models compare with status models in approaches to evaluating student achievement. Chapter II considers, for each of the nine pilot grantee states in the 2007–08 school year, the impact of the state's GMPP model on its AYP determinations. Chapter III examines the effect of the growth models on AYP outcomes in different types of schools.

The implications of the initial pilot project experiences for future efforts to use growth models are developed in Chapter IV. Analyses in that chapter address a number of hypothetical questions about how results might change if the data collected as part of the pilot project were used differently. This section also considers how the type of growth model selected affected on-track determinations and how the required focus on grade-level proficiency affects AYP outcomes. This section also evaluates the adequacy of state longitudinal data systems for implementing growth models.

Some of the analyses in Chapter IV address hypothetical questions about alternatives to current *ESEA* regulations for the Growth Model Pilot Project and would thus only become practical if the regulations were changed. While these analyses are part of the contractual scope of work for the evaluation project reported on here, they do not reflect any type of an endorsement by ED of the alternatives analyzed.

[5] See "Secretary Spellings Invites Eligible States to Submit Innovative Models for Expanded Growth Model Pilot" (retrieved September 2010 from http://www2.ed.gov/news/pressreleases/2007/12/12072007.html) and "U.S. Secretary of Education Margaret Spellings Approves Additional Growth Model Pilots for 2007–08 School Year" (retrieved June 2008 from http://www.ed.gov/news/pressreleases/2008/06/06102008.html) and "Secretary Spellings Approves Additional Growth Model Pilots for 2008–09 School Year" retrieved January 2009 from http://www.ed.gov/news/pressreleases/2009/01/01082009a.html).

The Status Model of Accountability Under *ESEA*

Under *ESEA* as amended in 2001, each state revised its standards-based system of student achievement measures and targets and began conducting annual assessments of proficiency levels to determine whether its schools and districts were making AYP.[6] Each school has a certain percentage of students who score proficient or higher each year on the mathematics and reading or language arts achievement tests, and this constitutes an annual measure of a school's performance. That percentage is expected to reach 100 percent by the end of the 2013–14 school year in incremental steps. In addition, *ESEA* requires each school to meet or exceed statewide standards on one or more "other academic indicators," typically defined in terms of average daily attendance for elementary schools and graduation rate for high schools.

Each step in the path to achieving universal proficiency in reading and mathematics under *ESEA* is known as the "annual measurable objective" or "AMO."[7] The AMO is the standard that schools and districts use to determine whether or not they are making AYP. AMO trajectories are not uniform across states, including the states in this study. Some increase in a consistent linear fashion toward 2014, while others increase more in the years closer to 2014 than in those closer to the 2002 start of the AYP requirements.

In order for the school to make AYP under the *ESEA* status model, several conditions must be met. These conditions are required by the law and are intended to improve the reliability and validity of the accountability results. First, the school must test at least 95 percent of its students in each *ESEA* reporting group in both reading or language arts and mathematics. These *ESEA* reporting groups consist of all students plus major racial and ethnic subgroups, students with disabilities, limited English proficient students, and students from low-income households. Within each school, a reporting group may be excluded from federal accountability requirements if the number of "full academic year" students from that group is below a minimum "*n*" size. Most states define a full academic year as starting in the fall when enrollments are finalized (typically around Oct. 1) and extending through the end of the testing window in the spring, while the state-defined minimum *n* sizes for reporting groups range from a low of no minimum to 100 students. Second, the percentage of tested students scoring proficient or higher must meet or exceed the AMO in both subjects for eligible reporting groups. These percentages are calculated only for students enrolled for the full academic year. If a single subgroup fails to achieve the AMO, the school does not make AYP.

In order to reduce the chances of incorrectly classifying schools as not making AYP, the states are allowed to apply additional steps within the status model. If any one or more of the *ESEA* reporting groups did not make AYP, the school may:

---

[6] See *Elementary and Secondary Education Act*, Title I – Improving the Academic Achievement of the Disadvantaged, Section 1111, paragraph (b)(2) retrieved October 2010 from http://www2.ed.gov/policy/elsec/leg/esea02/pg2.html#sec1111 .

[7] AMO trajectories are defined by the states. They have levels for each school year that increase at least every three years and can be different within each year across grades and subjects. States were required to follow strict statutory requirements in setting their initial AMOs.

- Apply a confidence interval to the group's percent proficient and compare the upper bound to the AMO. Analogous to the margin of error typically reported with results from political and other opinion polls, a confidence interval represents the range of values within which the true value is expected to fall for a given level of statistical certainty (e.g., 95 percent). The higher the standard of certainty asked for, the wider the confidence interval. For a given standard of certainty, the confidence interval narrows as the student count grows. If the higher limit of that confidence interval is greater than the AMO, then the subgroup is considered to make AYP.
- Average the test results for the group over two or three years and compare the average to the AMO (this is often referred to as "multiyear averaging").
- Apply safe-harbor, whereby the group makes AYP if the percentage of non-proficient students in the group decreased by 10 percent or more from the prior year (e.g., a decline from 30 percent to 27 percent non-proficient).
- Apply safe-harbor but assess whether the reduction in the percentage non-proficient was 10 percent or more from the average percentage non-proficient over the prior two or three years.

Variations of this basic method for determining school AYP are used by all states and are collectively referred to as the status model. A simplified version of the decision tree (excluding the full academic year, minimum *n*, confidence intervals, and multiyear averaging conditions) is illustrated in Exhibit 1.

Exhibit 1
Determining AYP Under the Status Model

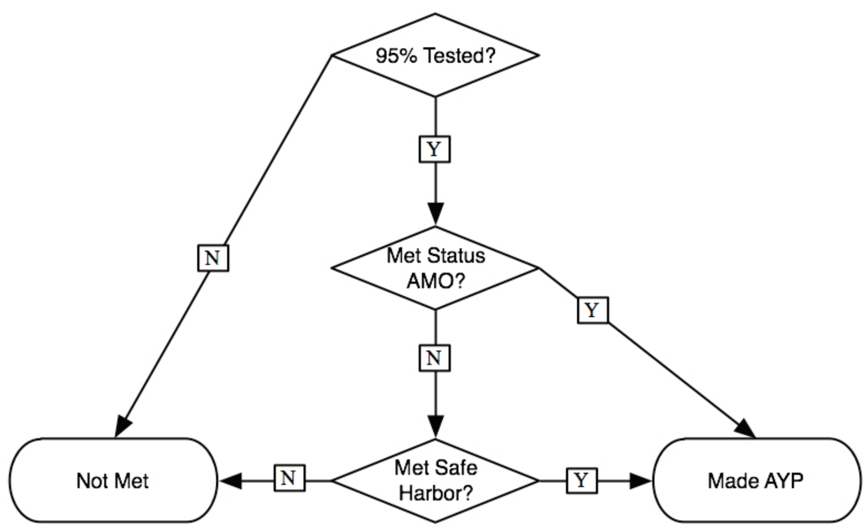


Exhibit reads: A school's AYP under the status model is determined by first assessing whether 95 percent of the students in each reporting group took the reading and the mathematics tests. If so, the percentages of the test-takers that scored at or above the proficiency cut scores are calculated and are compared to the AMO. If the percentage proficient for any reporting group is below the AMO, then the group's percentage non-proficient is compared to its percentage non-proficient from the prior school year. If the current percentage non-proficient is 10 percent or more lower than the prior year, the group makes AYP by safe-harbor.

Limitations of the Status Model

One characteristic of the status model is that it does not recognize real improvements in student achievement unless they result in higher percentages of students meeting or exceeding proficiency standards in a given year. Schools with many students improving but still falling short of the proficiency standard may not meet their AMO and thus will not make AYP. Conversely, schools with many students at or above the proficiency standard will still make AYP even if few of these students improve from year-to-year.

The fact that schools are evaluated strictly in terms of the numbers of proficient students, with no credit given for improvement short of or over and above proficiency, has raised concerns that, in the short term, instructional resources may be focused on students who are closest to the proficiency threshold ("bubble students"). Students less likely to attain proficiency from a given amount of instructional effort—e.g., those farthest below the proficiency threshold—may receive less attention. Studies investigating this hypothesis have come to mixed conclusions, however, and rising AMOs make this strategy less relevant because all students must be proficient by 2014 in order for a school to make AYP. On the other side of the ledger, high-scoring students may receive less attention because there are no statutory consequences for failing to improve achievement among those already proficient.[8] This possible tendency is tied to the use of a minimum proficiency threshold and would thus not be affected by the 2014 target.

Another characteristic of the status model is that it does not take account of changes in a school's student composition from one year to the next. Thus a school classified as not making AYP in one year could be judged to make AYP in the next if more proficient students enrolled or if less-proficient students left. Conversely, a school making AYP one year may not reach the AMO standard the following year if its student composition shifted the other way.

The Growth Model Pilot Project (GMPP)

The U.S. Department of Education (ED) initiated the Growth Model Pilot Project (GMPP) in November 2005 with the goal of approving up to ten states to incorporate growth models in school AYP determinations under *ESEA*. Growth models are defined as complements or alternatives to the standard status model for determining school AYP; they base AYP on some measure of how much students have gained from one year to the next.

Growth models are intended to recognize schools' progress moving students toward proficiency. As suggested above, student growth patterns in a school may overlap with or diverge from

---

[8] Research on the extent to which schools have adopted these sorts of strategies has been conducted by Naomi Chudowsky, Victor Chudowsky, and Nancy Kober (2009), "Is the Emphasis on 'Proficiency' Shortchanging Higher- and Lower-Achieving Students?" Retrieved July 2009 from http://www.cep-dc.org/index.cfm?fuseaction=document_ext.showDocumentByID&nodeID=1&DocumentID=280; Derek Neal and Diane Whitmore Schantzenbach (August 2007), NBER Working Paper 13239 "*Left Behind By Design: Proficiency Counts and Test-Based Accountability*" retrieved September 2008 from http://www.nber.org/papers/w13293.pdf ; and Jennifer Booher-Jennings (2005), "Below the Bubble: 'Educational Triage' and the Texas Accountability System," *American Educational Research Journal*, 42: 231–268.

assessments of students' proficiency based on the status model; these are illustrated in Exhibit 2. Ideally, all schools would be in cell A, meeting the status requirements and realizing high rates of annual growth. The basic goal of the GMPP is to identify schools in cell B, that is, schools with high numbers of students making progress but not yet attaining the grade-level proficiency thresholds necessary to meet AYP standards. Schools in cell C—those with low rates of student growth but still making AYP—are also of interest to the GMPP but, as will be discussed further at various points in this report, were not targeted in the project. The overarching goal of *ESEA* as implemented by the GMPP is to ensure that no schools are in cell D, not meeting the status model requirements and not making sufficient gains to be on-track to meet the requirements in the near future.

Exhibit 2
Relationship of Growth and Status Models for Assessing Achievement Proficiency

	School AYP Designation Under the Status Model	
School-level Growth in Achievement	**Made AYP**	**Did Not Make AYP**
High rates of growth	Cell A	Cell B
Low rates of growth	Cell C	Cell D

The U.S. Department of Education used a rigorous peer review process to evaluate the adequacy of the technical aspects of the proposed models and to ensure that the models aligned with seven core principles.[9] These core principles required all pilot states to set annual "growth targets" for ensuring universal grade-level proficiency by 2014 and to track individual students across schools and measure their progress across grades in both reading and mathematics. The first principle requires that the growth model, like the status model, be applied to each targeted subgroup as well as all students in the school. This means that growth outcomes are to be monitored separately, or "disaggregated," for major racial and ethnic groups, limited English proficient (LEP) students, special education students, and low-income students.

The second principle stipulates that growth expectations cannot be based on student background or school characteristics. This principle is consistent with the *ESEA* rule that proficiency targets must be the same for all students in a given grade and cannot be modified for different kinds of students or schools. In the growth model context, this excludes use of the "relative-growth" models permitted under the 1994 amendments that gave AYP credit to schools or subgroups strictly on the basis of realizing average or even higher-than-average annual growth rates. The key criterion under the 2001 reauthorization of *ESEA* is "meeting grade-level proficiency." Students scoring below proficiency must not only gain more per year than one grade-level equivalency, but those gains must also point to attaining proficiency standards within a specified

[9] See U.S. Department of Education (Nov. 18, 2005) "Press Release: Secretary Spellings Announces Growth Model Pilot, Addresses Chief State School Officers' Annual Policy Forum in Richmond" (retrieved May 2008 from http://www.ed.gov/news/pressreleases/2005/11/11182005.html), and U.S. Department of Education (July 2007) "No Child Left Behind fact sheet: Growth Models—Ensuring Grade-Level Proficiency for All Students by 2014" (retrieved May 2008 from http://www.ed.gov/admins/lead/account/growthmodel/proficiency.pdf).

time frame.[10] A full list of the seven core principles of the Growth Model Pilot Project is provided in Exhibit 3.

Exhibit 3
Seven Core Principles of the Growth Model Pilot Project

States approved for participation in the GMPP were required to meet seven core principles in the *ESEA* accountability plans they submitted for incorporating growth models in their AYP measurements:

1. Ensure that all students are proficient by 2014, and set annual goals to ensure that the achievement gap is closing for all groups of students;
2. Set expectations for annual achievement based on meeting grade-level proficiency, not on student background or school characteristics;
3. Hold schools accountable for student achievement in reading or language arts and mathematics;
4. Ensure that all students in tested grades are included in the assessment and accountability system, hold schools and districts accountable for the performance of each student subgroup, and include all schools and districts;
5. Include assessments in each of grades 3–8 and in high school for both reading or language arts and mathematics, and ensure that they have been operational for more than one year and receive approval through the *NCLB* peer review process. The assessment system must also produce comparable results from grade to grade and year to year;
6. Track student progress as part of the state data system; and
7. Include student participation rates and student achievement on a separate academic indicator in the state accountability system.

Source: See "Peer Review Guidance for the *NCLB* Growth Model Pilot Applications," January 25, 2006 (retrieved May 2008 from http://www.ed.gov/policy/elsec/guid/growthmodelguidance.pdf).

Other significant features of pilot growth models were allowed to vary, as long as the technical specifications passed review by a panel of nationally recognized experts.[11] Reviewers had a series of meetings to discuss the Peer Review Guidance document and the Department's expectations for the process. Proposals that passed an initial review by Department staff and a round of clarification by states were forwarded to the peer reviewers. The Department then set up conference calls between the states and the peer review team, after which the reviewers met again to discuss the proposals and make recommendations for use by the secretary in deciding which proposals to approve for the pilot project.

[10] This guiding principle also precludes the use of growth models referred to as "value added" models (VAM), at least when they include statistical adjustments for student social background variables or school characteristics. In general terms, value-added formulations seek to separate the effects of various factors on student growth in order to make attributions about the marginal effectiveness of the factors on achievement gains. Inputs are variously defined, depending on the purpose of the analysis, but often include student social background variables, teachers, and instructional programs. If used for evaluation of schools, value-added models estimate growth for a common "average type" of student in each school and the schools are then compared in terms of that background-standardized estimate. In that sense, VAM models are typically measures of *relative school effectiveness*, not student proficiency or growth toward proficiency. The latter is the purpose of GMPP, and VAM are generally inappropriate for that purpose.

[11] See "Peer Review Guidance for the *NCLB* Growth Model Pilot Applications," January 25, 2006 (retrieved May 2008 from http://www.ed.gov/policy/elsec/guid/growthmodelguidance.pdf).

Of the 20 states that submitted proposals by February 2006, 13 asked for approval to use growth models for immediate use in the 2005–06 school year while the others proposed to start in 2006–07. Eight of these proposals passed the initial evaluation, and revised proposals were forwarded to the peer reviewers.[12] Two states, North Carolina and Tennessee, received approval to use growth models for 2005–06, while Delaware, Arkansas, and Florida revised their proposals and were approved in a subsequent peer review to implement their proposed models for the following school year (2006–07). A second round of proposals was solicited in October 2006. Six previous applicants submitted revised proposals along with submissions from three new states.[13] Of these, Alaska, Arizona, and Iowa received immediate or conditional approval to use their growth models beginning with the 2006–07 school year, while Ohio was approved to begin in the 2007–08 school year.[14]

The approved pilot states were all required to calculate whether each student in each target grade level was on-track to be proficient within a specified number of years or by a particular grade level. The growth models determined whether each student was on-track to reach or exceed the state's grade-level proficiency cut scores on the annual tests of reading and mathematics within three or four years or by a specified grade level (usually grade eight or nine) as defined by the state's particular growth model. Students who were on-track to be proficient could be counted as proficient for the purposes of determining AYP. However, the states were given a variety of options on how to incorporate the student growth indicator in their AYP determinations.[15] These included using:

- only the growth measure to calculate the percentage on-track to proficiency for AMO assessment;
- both the status and the growth measures to calculate the percentage proficient or on-track for AMO assessment;

[12] These reviewers included Eric Hanushek, Stanford University, Chris Schatschneider, Florida State University, David Francis, University of Houston, Margaret Goertz, University of Pennsylvania, Kati Haycock, The Education Trust, William Taylor, Citizens Commission on Civil Rights, Sharon Lewis, Council of Great City Schools (retired), Robert Mendro, Dallas Independent School District, Jeff Nellhaus, Massachusetts Department of Education, and Mitchell Chester, then at the Ohio Department of Education.

[13] See "Secretary Spellings Approves Additional Growth Model Pilots for 2006–2007" (retrieved June 2008 from http://www.ed.gov/news/pressreleases/2006/11/11092006a.html). This panel of reviewers included Anthony Bryk, Stanford University, Harold Doran, American Institutes for Research, Chrys Dougherty, National Center for Educational Accountability, Lou Fabrizio, North Carolina Department of Public Instruction, Tom Fisher, Independent Consultant, Pete Goldschmidt, University of California at Los Angeles, Sharon Lewis, Council of Great City Schools (retired), Margaret McLaughlin, University of Maryland, Robert Mendro, Dallas Independent School District, Jeff Nellhaus, Massachusetts Department of Education, Ann O'Connell, Ohio State University, Dianne Piché, Citizens Commission on Civil Rights, Sandy Sanford, Riverside County Office of Education, Chris Schatschneider, Florida State University, William Taylor, Citizens Commission on Civil Rights, and Martha Thurlow, University of Minnesota.

[14] In December 2007, the Department moved to expand participation in the growth model project and extended an invitation to all states to apply for inclusion. At the time of writing, 15 states have been approved to implement growth models in AYP determinations.

[15] See "Peer Review Guidance for the *NCLB* Growth Model Pilot Applications" (retrieved May 2008 from http://www.ed.gov/policy/elsec/guid/growthmodelguidance.pdf).

- the status measure to calculate the percentage proficient, applying safe-harbor provisions if needed, and using the growth measure either in conjunction with the status measure or alone if AYP was not met with safe-harbor; and
- safe-harbor provisions and the growth measure for AMO assessment.

While the approved models differ from one another in a number of important ways, all use state-specific assessment data to measure student progress and proficiency, and the method of incorporating growth outcomes in AYP determinations was generally the same. Eight of the nine pilot states proposed to apply growth criteria only *after* schools failed to make AYP under the status and safe-harbor provisions rather than determining AYP solely on the basis of student growth. More specifically, the designs of the pilots in these eight states applied growth criteria only to those students who were members of *ESEA* reporting groups that did not reach their AMOs or attain AYP via safe-harbor provisions. This basic method of augmenting status with growth results is shown in Exhibit 4 (again, this is a simplification in that provisions for using confidence intervals and multiyear averaging prior to applying growth results are not represented).

Exhibit 4
Determining AYP Under Status, Safe-Harbor, and Growth

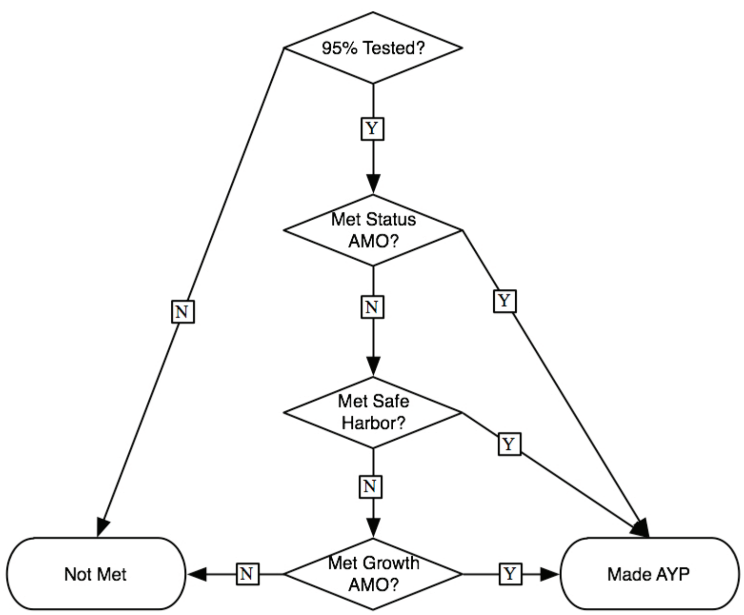


Exhibit reads: School AYP under the status-plus-growth model is determined by first assessing whether 95 percent of the students in each reporting group took the reading and the mathematics tests. If so, the percentages of the test-takers that scored at or above the proficiency cut scores are calculated and are compared to the AMO. If the percentage proficient for any reporting group is below the AMO, then the group's percentage non-proficient is compared to its percentage non-proficient from the prior school year. If the current percentage is 10 percent or more lower than the prior year, the group makes AYP by safe-harbor. If any of the groups that failed to meet the status AMO also failed safe-harbor, then, in most of the pilot states, the number of non-proficient students who are on-track to proficiency per the growth model are added to the number of proficient students and the percentage proficient or on-track compared to the AMO. If one or more groups meets or exceeds the AMO because of the addition of the on-track students, the school is classified as making AYP by growth.

Delaware adopted a different order, applying growth first and then applying status and safe-harbor only to schools and subgroups that did not meet AYP under the growth criteria. This procedure thus classified schools as making AYP by growth even if they also would have made AYP by status or safe-harbor criteria. In contrast, the schools identified as making AYP by growth in the other states were all cases in which AYP was not met by status or safe-harbor.

Types of Growth Models Implemented in the Pilot

The states approved for the pilot study proposed to employ either a status-augmented-with-growth or (in Delaware) a growth-augmented-with-status method of determining AYP, but the models used by the pilot states varied in how they established growth targets that define whether individual students were "on-track" to reach proficiency in the allotted time frame. The pilot states used three basic types of growth models: the *transition matrix* model, the *trajectory* model, and the *projection* model.[16]

Transition Matrix. This type of model evaluates student progress from year to year in terms of a small set of discrete performance levels. The levels are defined in general terms that are applied to all grades (e.g., below proficient, proficient, advanced). Student growth is indexed by movement ("transitions") from lower to higher categories. Delaware and Iowa used models of this type.

An illustration of the transition matrix model used in Iowa is shown in Exhibit 5. Students who scored below the proficiency threshold in year 1 (the first three rows) were classified as "on-track" to proficiency if their test performances improved enough in year 2 to move up at least one level. All students who were proficient or advanced in year 1 and who were proficient or advanced in year 2 were classified as on-track. For the purpose of determining AYP, the Iowa model counted students who were on-track toward proficiency as being fully proficient. All students in the gray-shaded cells are not on-track and do not count as proficient for AYP purposes.

Exhibit 5
Illustration of How the Iowa Transition Matrix Model Classifies Students

Year 1 Performance Level	Year 2 Performance Level				
	Weak	Low Marginal	High Marginal	Proficient	Advanced
Weak	Off-track	On-track	On-track	On-track	On-track
Low marginal	Off-track	Off-track	On-track	On-track	On-track
High marginal	Off-track	Off-track	Off-track	On-track	On-track
Proficient	Off-track	Off-track	Off-track	On-track	On-track
Advanced	Off-track	Off-track	Off-track	On-track	On-track

[16] See the CCSSO's *Implementer's Guide to Growth Models* for an alternative, more extensive, typology of growth models (retrieved May 2008 from http://www.ccsso.org/content/pdfs/IGG%20Final%20AP.pdf).

Delaware used a somewhat different type of transition matrix model; it is illustrated in Exhibit 6. It used a point system that gave a maximum of 300 points to all students who attained proficiency in year 2 but gave only partial credit to students who scored below proficiency in year 1 and who made gains but did not reach the proficiency standard in year 2. Students who scored below the proficiency threshold in year 1 and did not move up to a sufficiently higher level in year 2 (the gray-shaded cells) were assigned no points for determining AYP. Points were summed within subgroups and divided by the number of students in each subgroup to calculate an "average growth value" which was compared to annual growth targets.

Exhibit 6
Illustration of How the Delaware Transition Matrix Model Assigns Points to Students' Gains

Year 1 Performance Level	Year 2 Performance Level				
	PL 1A	PL 1B	PL 2A	PL 2B	PL 3, 4, and 5
PL 1A: lowest level below proficiency	0	150	225	250	300
PL 1B	0	0	175	225	300
PL 2A	0	0	0	200	300
PL 2B: highest level below proficiency	0	0	0	0	300
PL 3, 4, and 5: proficient or higher	0	0	0	0	300

Trajectory. This type of model uses the gap between a student's baseline test score and a performance standard several years out to calculate the amount of growth that student must attain to become proficient. This performance "trajectory" is divided up into annual growth targets that, taken together, put the student on-track to reach a grade-level proficiency cut score within the allotted years. Alaska, Arizona, Arkansas, Florida, and North Carolina used trajectory models.

An example of a trajectory model is shown in Exhibit 7. The Y axis represents student achievement on a vertically aligned scale, and the X axis represents successive grades. The large solid dots indicate the level of achievement necessary to be considered proficient at each time point ("p1" for proficiency at time 1, "p2" for proficiency at time 2, and "p3" for proficiency at time 3). The hollow dots represent a student's actual achievement at time points 1 and 2, marked "s1" and "s2" respectively.

Exhibit 7
Illustration of How Trajectory Models Set Targets for Students Scoring Below Proficiency Thresholds

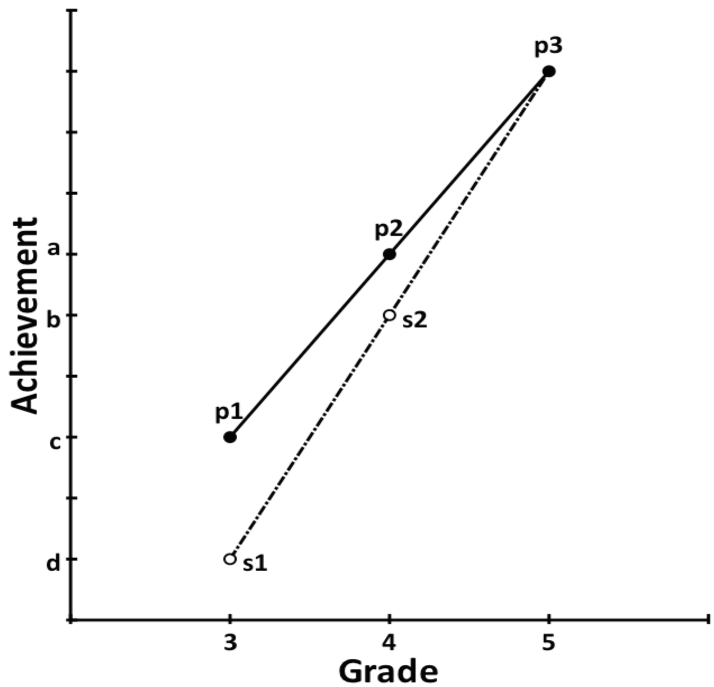


Exhibit reads: Under a trajectory model with a fifth-grade time limit, a student with a third grade score s1 below the proficiency cut score p1 must score at least as high as s2 in the fourth grade in order to be on-track to proficiency by fifth grade.

At grade 3, the student is not considered proficient because his or her achievement ("s1") is lower than the proficiency cut score at that time ("p1"). Under a trajectory model, each student has a growth target that must be met to be considered "on-track" and count as proficient. In the illustration above, this is done by drawing a line from the student's achievement at grade 3 ("s1") to the proficiency point for grade 5 ("p3"). This line intersects point "s2" at grade 4, indicating that "s2" is what the student must achieve in grade 4 to be considered "on-track" to proficiency. The level of achievement targeted by the trajectory model is lower than what is expected under the status model. Under the status model, a student must have a level of achievement at or above point "p2."

This illustration shows that trajectory models allow students to score below the proficiency thresholds between the year they first miss proficiency ("s1") and the year when they must be proficient under status criteria ("p3"). Because expected growth is based on a trajectory from the student's initial achievement and the target year, trajectory models distribute the growth required to meet status expectations among the intervening years. While this illustration shows a linear set of proficiency cut scores and a corresponding linear trajectory for being on-track, some states, such as Arkansas, have a nonlinear path of cut scores and use a nonlinear method to calculate growth targets.

Projection. This type of model uses current and past test scores to statistically predict performance several years ahead based on how all students in the state or school with similar patterns of scores generally perform. Such "projections" use multiple test scores from a reference cohort of students to estimate prediction equations using multiple regression techniques. For example, the eighth-grade cohort of students in 2007–08 could be the reference cohort, and the regression would estimate equations relating all of the past test scores for these students to their grade 8 reading and mathematics scores. These equations are then used to predict how each student in younger cohorts (e.g., the seventh-grade cohort in 2007–08) will score at the end of the eighth grade, given their current and prior test scores. If that predicted score is equal to or greater than the proficiency cut point, the student is classified as on-track to proficiency. Ohio and Tennessee are using projection models.

Exhibit 8 illustrates a simple projection model. The projection model determines a projected level of achievement for a future time point for each student. Depending on the specifics of a state's model, students who do not make the proficiency threshold under a status model at grade 5 may still be counted as proficient if the projection model predicts that they will make proficiency by grade 6. Like the trajectory model illustration, the Y axis represents student achievement on some vertical scale. The X axis represents grades. Again, the solid dots represent grade-level proficiency thresholds at grades 3 (p1) through 6 (p4), and the hollow dots represent a student's achievement over consecutive grades. The solid line represents the projection equation based on the current and past data for that student (points "s1" through "s3"). Though illustrated here with data for just one student, both the Ohio and Tennessee models estimate this projection equation from a regression analysis of students from a standard-setting cohort that most recently completed the eighth grade, and then use the regression coefficients to predict current students' future scores based on their prior scores. The dotted line is the extension of that projection equation to a future time point specified by the state's growth model.

Under a status model, a student who reached the level of "s3" would not be considered proficient at grade 5, because his or her score is lower than that of point "p3". However, under a growth model using projections based on student data, that student would be considered "on-track" because his or her projected score meets or exceeds the proficiency cut score at grade 6 (the dotted line is above "p4" in the figure).

Exhibit 8
Illustration of How a Projection Model Predicts Future Achievement Based on Prior Scores

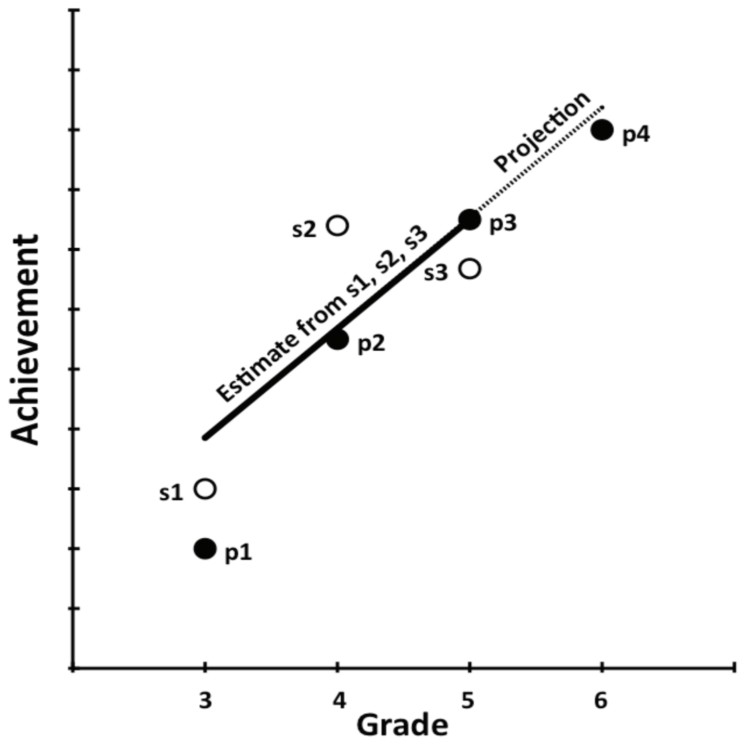


Exhibit reads: Under the projection model, a regression line is estimated on the basis of past scores s1, s2, and s3 to predict sixth-grade achievement.

How Growth Model Results Are Used for School AYP Determinations

The pilot states used the student-level indicators of on-track-to-proficiency generated by the GMPP growth models to augment the status indicators of proficiency in order to make AYP determinations for their schools. Eight of the nine states with data available for 2007–08 followed what is referred to in the rest of the report as a “status-plus-growth” methodology. This involved using the standard status model first to assess whether the school made AYP. If the school did not make AYP by status, safe-harbor criteria (i.e., a 10 percent reduction in the number of non-proficient students from the prior to the current year) were applied. If the school still did not make AYP, only then were the growth model results used.

An important aspect of the AYP determinations is that they are rooted in reading or language arts and mathematics achievement outcomes for all grade-eligible students in the school and also for each official reporting subgroup of students. Before the achievement scores are evaluated, an initial screening is done to assess whether the enrollment numbers and participation rates are sufficient. In terms of enrollment numbers, AYP determinations must be made for all schools regardless of size. However, each state identifies a minimum number of students that must be enrolled in order for a subgroup to be included in AYP calculations; in most states, this "minimum n" is 30 or 40 students. If a school does not enroll enough students in a subgroup, the students in that subgroup are counted in the "all students" category, which every school must define. For the enrolled students, the federal law requires that a minimum of 95 percent must take each test, and this rate must be realized for all students and for each subgroup reaching the minimum n size for participation.[17] If one or more eligible subgroup did not attain 95 percent participation, the school fails to make AYP regardless of test score results.

In the eight status-plus-growth pilot states, a school was designated as "made AYP by status" if all students and all subgroups had sufficient numbers of students scoring at or above the proficiency level in both subjects to meet their annual measurable objective (AMO, usually defined as a percentage of students at or above proficiency). If one or more subgroups failed to meet one or both AMOs, then the multiyear averaging and confidence interval methods described on p. 3 would be applied to the subgroup(s), still under the auspices of "status." If one or more subgroups still did not meet an AMO, then the safe-harbor method was applied to the subgroup(s) and, if the group(s) then met the criterion of a 10 percent reduction in the number of non-proficient students, the school was designated as "made AYP by safe-harbor."

Only at the end of the status and safe-harbor tests were the growth model on-track indicators used in the status-plus-growth states. If one or more subgroups still did not pass after status and safe-harbor, six of the pilot states added the non-proficient students who were on-track according to the growth model to the proficient students and calculated the percent "proficient or on-track." In two states, the on-track indicators were used for all students in the subgroups that had not passed by status or safe-harbor. In either case, if that new percent met or exceeded the AMO, the group was classified as "made AYP by growth;" if all of the subgroups that failed to make AYP by status or safe-harbor passed when the on-track students were added in, then the school as a whole was classified as "made AYP by growth." In the states that assess AMOs by adding on-track students to those already proficient, it is thus possible for a schoolwide designation of "made AYP by growth" to result from just *one* non-proficient student being on-track to raise the percentage of "proficient plus on-track" students enough to meet the AMO. In general, then, "making AYP under growth" does *not* necessarily mean that all or even most students in the school are making sufficient learning gains to be counted as on-track to reach or (if already reached) maintain proficiency.

In contrast to the other eight pilot states, Delaware used a "growth-plus-status" method for determining AYP. As discussed more fully in Chapter II and Appendix B, Delaware used a point system whereby all students scoring proficient or higher earned a maximum number of points and students scoring below the proficiency cut score earned some fraction of that maximum depending on how much they improved from the prior year. If the average of the proficient (full credit) and non-proficient (partial credit) students' points met or exceeded an annual target AMO, then the subgroup was classified as making AYP by growth in that subject. If all students and subgroups met the AMO, then the school was classified as making AYP by growth. If one or more subgroup did not meet the AMO, then multiyear averaging and

[17] States are allowed to set their minimum n and the numbers range from no minimum to 100 in California.

confidence intervals were applied to their point averages to see if they overlapped with the target AMO. If not, then safe-harbor was applied.

Exhibit 9 provides an overview of each pilot state's growth model organized by the type of model used and provides brief model summaries along dimensions that may affect final AYP outcomes. These include which grades receive growth calculations, how many years of growth are allowed, which tests are used to measure achievement, how students are identified as on-track per the growth model, and how the on-track data are used for school AYP determinations. More detailed information about the technical features of each model can be found in the growth model summaries in Chapter II and also in Appendix B.

Exhibit 9
Overview of Growth Models Approved for the GMPP in the 2007–08 School Year

State	First Year Used	Grades Included	Minimum Subgroup Size	Years to Proficient	Achievement Measures Used	How is Student On-Track-to-Proficiency Determined?	How are On-Track Determinations Used for AYP Determinations?
Delaware	2006–07	3–10	40	N/A	Scale scores for reading and math via the Delaware Student Testing Program	*Transition matrix model:* Four levels of "below the standard" are used to categorize non-proficient students. A point system awards full credit for proficiency and partial credit for movement to higher levels of below proficient.	Students' points are averaged to make AYP determinations for each subgroup. If the subgroup's average meets or exceeds the AMO, the subgroup makes AYP.
Iowa	2006–07	4–8	30	4	Scale scores based on biannual Iowa Test of Basic Skills (ITBS) math and reading results	*Transition matrix model:* Three categories of performance are used to classify non-proficient students. Non-proficient students who move to a higher category of non-proficiency are on-track.	Status and safe-harbor applied first; if a subgroup does not make AYP with either, on-track students are added to proficient students.
Alaska	2006–07	4–9	25	4	Adjusted scale scores for math and language arts (reading + writing) on the Standards Based Assessment (SBA) Test	*Trajectory model:* Annual growth targets for each student are based on the test score gains needed to maintain or reach proficiency in four years or by grade 10. Students who close the test score gap by the reciprocal of the years remaining (e.g., 1/4, 1/3, 1/2) are "on-track."	Status and safe-harbor applied first; if a subgroup does not make AYP with either, on-track students are added to proficient students.
Arizona	2006–07	4–7	40	3	Regression-adjusted scale scores for math and reading on the Arizona Instrument to Measure Standards (AIMS) tests	*Trajectory model: :* Annual growth targets for each student are based on the test score gains needed to maintain or reach proficiency in three years or by grade 8. Students who close the test score gap by the reciprocal of the years remaining (e.g., 1/3, 1/2) are "on-track."	Status and safe-harbor applied first; if a subgroup does not make AYP with either, on-track students are added to proficient students.

Exhibit 9 continued next page

Exhibit 9 (continued from previous page)

State	First Year Used	Grades Included	Minimum Subgroup Size	Years to Proficient	Achievement Measures Used	How is Student On-Track-to-Proficiency Determined?	How are On-Track Determinations Used for AYP Determinations?
Arkansas	2006–07	4–7	40	4	Scale scores for literacy and math on the Arkansas Benchmark Exams	*Trajectory model:* Annual growth targets for each student are based on the test score gains needed to maintain or reach proficiency in four years or by grade 8. Students who close the test score gap by a grade-specific amount are "on-track."	Status and safe-harbor applied first; if a subgroup does not make AYP with either, on-track students are added to proficient students.
Florida	2006–07	3–10	30	3	Developmental Scale Scores (DSS) for math and reading on the Florida Comprehensive Assessment Test	*Trajectory model:* Growth targets are based on the gap between initial score and proficiency cut scores three years later. Non-proficient students who close the original gap by one-third (one-half for ninth-graders) annually are on-track.	Status and safe-harbor applied first; if a subgroup does not make AYP with either, the percentage on-track in the subgroup is compared to the AMO.
North Carolina	2005–06	3–7	40	4	Scale scores on fall third-grade pretest and annual North Carolina End-of-Grade Math and Reading Tests	*Trajectory model:* Growth targets are based on closing the gap between baseline test score and proficiency cutoff in four years or by grade 8. Students who close the test score gap by 1/4 each year are on-track.	Status and safe-harbor applied first; if a subgroup does not make AYP with either, on-track students are added to proficient students.
Ohio	2007–08	4–7	30	3	Scores for reading and math on the Ohio Achievement Tests (plus the Ohio Proficiency Tests for the first cohort)	*Projection model*: Predicted scores for the grade beyond the school's terminal grade or the grade three years later are calculated for each student based on current and prior test scores; if predicted score plus two SE units is above the cut score for the target grade, student is on-track.	Status and safe-harbor applied first; if a subgroup does not make AYP with either, on-track students are added to proficient students.
Tennessee	2005–06	4–8	45	4	Scale scores for math and reading/language arts from the Tennessee Comprehensive Assessment Program (TCAP)	*Projection model:* Predicted scores for grade 8 are calculated for each student based on current and prior test scores; if predicted score is above the cut score for the target grade, student is on-track.	Status and safe-harbor applied first; if a subgroup does not make AYP with either, the percentage on-track in the subgroup is compared to the AMO.

Data Sources and Availability

Data employed for the analyses in this report were provided by the U.S. Department of Education and the pilot grantee states. The data from the Department were extracted from the ED*Facts* database. ED*Facts* is the main repository for school, district, and state data on *ESEA* requirements related to making AYP determinations, as well as enrollment and demographic data. All states are required to submit standard data elements on all their districts and K–12 schools each year to ED*Facts*. For the 2007–08 school year data that are the focus of this report, the standard reporting variable on whether the school made AYP was modified to collect information on whether each school made AYP because of the growth model. This variable was defined in ED*Facts* with a set of mutually exclusive categories: "made AYP by regular determination," "made AYP by growth," or "did not make AYP."

The "made AYP by regular determination" category included all methods of making AYP except by growth criteria. Regular determination included status criteria alone, status with confidence intervals, status with multiyear averaging, and safe-harbor (see p. 3 for definitions of these terms). Of these, only safe-harbor was reported separately. The standard ED*Facts* reporting variable for AMO results in 2006–07 included the mutually exclusive categories of "met by status," "met by safe-harbor," "exempt by minimum n," and "did not meet" for each subject area. A new reporting variable for "met by growth" was added to subgroup AMO results in 2007–08. For purposes of this report, schools were classified as "made AYP by safe-harbor" if the school as a whole was identified as "made AYP by regular determination" and at least one subgroup was classified as "met by safe-harbor" for either the reading or mathematics AMO.

The nine GMPP states were additionally required as a condition for continued participation in the pilot to collect and make available for evaluation purposes student-level data on growth model outcomes. The data obtained directly from the pilot states consisted of scale scores and proficiency designations in reading or language arts and mathematics for each student in the grades involved in the GMPP. Of particular importance to this study were the proficiency designations, for these included the indicator of whether the student was "on-track" to achieve proficiency within the time frame specified by the approved growth model. Additional data elements included various background characteristics including school identifier codes, grade level of current enrollment, and *ESEA* subgroup memberships.

Eight of the nine states had provided both the ED*Facts* and student-level data required to address the two study questions which are the focus of this final report. The other state (Delaware) reported GMPP results to the ED*Facts* archive but, because of the unique way in which it used the growth model data, did not provide indicators of which schools classified as having made AYP by growth also would have made AYP if the growth data were not used. However, the Delaware Department of Education was able to provide that information separately from the ED*Facts* data.

Calculation of Hypothetical "Growth-Only" Outcomes

In addition to the ED*Facts*-reported outcomes for subgroups and schools, the student growth-model results are used in Chapter IV to calculate percentages of students who were on-track to reach proficiency. These "growth-only" indicators are used to address hypothetical "what if" questions that go beyond the basic descriptive aims of Chapters I–III.

As noted above, states had the options of making AYP determinations using (a) only growth results and (b) growth criteria before status and safe-harbor criteria. While these options were generally not exercised, it is possible to estimate what would happen if they had been. That information may prove useful to states in the process of developing growth models, as well as current pilot states contemplating changes to their models. Toward that end, the student indicators of on-track-to-proficiency are used to calculate hypothetical variants of AMO outcomes based on the percentages of on-track students in reading and mathematics for each *ESEA* subgroup in each school. These percentages are compared to the state's AMOs to assess the extent to which schools currently making AYP by status and safe-harbor would be able to reach their AMOs using growth-only percentages of students on-track to proficiency.

Data Limitations

The data described above have a number of limitations that are important to note at the outset and which will be reiterated at various points in this report. One limitation is that the mutually exclusive AYP categories reported by ED*Facts* made it impossible to determine the extent to which schools that made AYP under status and safe-harbor would also have made AYP if *only* growth criteria were considered. The student-level on-track-to-proficiency indicators described in the preceding subsection can be used to calculate growth-based AMO determinations for each *ESEA* reporting group and for the school as a whole. However, it should be emphasized that these growth-based AMO determinations are based solely on information regarding students' reading or language arts and mathematics performances and do not necessarily indicate whether a school would make AYP under a growth-only system. Under *ESEA*, schools must also meet additional conditions to make AYP. These conditions include the requirements that the school (a) meet or exceed a minimum level on an "other academic indicator," typically average daily attendance for elementary and middle schools and graduation rates for high schools; and (b) realize at least a 95 percent participation rate on the annual assessments.

A second limitation is that we were unable to consider these criteria in our calculations because the requisite data were not available. For example, the "other academic indicator data" are school-level determinations that were not provided in the student files and were unevenly reported in ED*Facts*. The student files from some states also did not include the information needed to calculate participation rates.[18] This means that some schools meeting the AMO on the

[18] Participation rates are calculated on the basis of students enrolled for the "full academic year" (FAY), but FAY and non-FAY students were not distinguished in some of the state data files.

achievement outcomes using the growth criteria identified in this report may have officially missed AYP on the additional criteria.[19]

Given the potential discrepancies between the growth-based AMO designations and the official *ESEA* designations, the results presented in Chapter IV of this report should be regarded as suggestive rather than definitive. With that caveat, these comparisons allow an assessment of the extent to which schools making AYP by status or safe-harbor criteria might also have made AYP if only growth criteria were used.

[19] The number of such "false positives" is probably small. National data from 2003–04 show that about 3 percent of all schools did not make AYP solely because of not making acceptable levels on the other academic indicator or low (below 95 percent) participation rates for the achievement testing (U.S. Department of Education, 2007, p. 43.)

II. Effects of Growth Models Implemented Under the GMPP on School AYP Determinations

This chapter presents overviews of the growth models implemented in the nine pilot states, focusing on the methods they employed to (1) set expectations, (2) measure growth, and (3) incorporate these growth results into their AYP determinations. The nine states provided data on the results of their pilot growth models and, following the overview of each state's model, we present evidence of the impact of their growth model on AYP outcomes.

The model summaries provided here are based on a review of extant documentation, including the approved GMPP proposals, decision letters from the U.S. Department of Education, and correspondence with the Department concerning feedback and suggested revisions from the peer review panels; growth model descriptions found in state accountability workbooks, school report cards, and other online technical documentation; and edits by state officials to draft model summaries provided to each pilot state at a December 2008 summit meeting in Washington, D.C. Any inconsistencies in model summaries were then resolved through e-mail correspondence and follow-up phone calls with state officials over the course of the evaluation. More detailed technical summaries of the states' growth models are included in Appendix B.

The data analyses address the research question "How many schools made AYP under the growth model that would not have made AYP under the *ESEA* status model?" For the eight states that participated in the GMPP for at least two years (Ohio joined the pilot project in 2007–08), we present data on the growth model outcomes in both the 2006–07 and 2007–08 school years. The data from both years are drawn primarily from the school-level reports from each state to the federal ED*Facts* system maintained by the U.S. Department of Education.

State Growth Models and Their Effects on Schoolwide AYP Results

Alaska's Growth Model

Alaska was formally accepted into the Growth Model Pilot Project on July 3, 2007. Alaska's growth model includes fourth- through ninth-graders only, with all other students expected to be proficient under the status model criteria. The state uses a trajectory model that defines targets for each student at each grade based on the student's scores on the Standards Based Assessment (SBA) tests for mathematics and language arts. The SBA test was first given in the 2004–05 school year to students in grades 3–9. The SBA is scaled such that students in every grade must score 300 or above to be considered proficient in math and a combined reading and writing score of 600 or above to be considered proficient in language arts.[20]

[20] Alaska estimates a "true score" for the SBA by adjusting for the annual reliability of the tests. This has the effect of raising scores that are below average. The estimated true score is calculated by subtracting an estimated "reliable deviation" of the student's observed score from the statewide average for students in the same grade.

Starting in the 2006–07 school year, fourth- through sixth-grade students scoring below the proficiency cutoff are counted as proficient for AYP purposes if they are on-track to reach proficiency by the seventh grade, and eighth- and ninth-graders are counted as proficient if they were on-track to reach proficiency by the tenth grade. Students are always only classified by the status model in grades 3 and 10. Annual growth targets for all fourth-, fifth-, sixth-, seventh-, eighth-, and ninth-graders who scored below proficiency are set by dividing the difference between the student's "baseline" test scores and the proficiency cutoffs (300 in math and 600 in language arts) by the number of years allowed. Students scoring below the proficiency cut score are classified as on-track to proficiency if they (1) meet or exceed their target score defined by the growth model, and (2) do not score lower than they did in the prior year.

The Alaska growth model only applies to students who score below proficient on the SBA in two or more consecutive years within the span of grades 3 through 8. The baseline scores used in the growth trajectory calculations are defined as those from the first year the student scored below the proficiency cutoff. For students in grades 4 through 6 in the 2007–08 school year, this could go back to the third grade (i.e., to 2004–05 when the SBA was first administered). Thus for fourth-grade students who scored below proficiency in both the third and fourth grades, the baseline is the third-grade score. For fourth-grade students who were at or above the proficiency cut score in third grade but below proficiency in fourth grade, the baseline is the fourth-grade score. For fifth-grade students who scored below proficiency in the third and fourth grades, their third-grade scores would also be the baselines used to define their trajectories to proficiency by grade 7. The baseline for students in grades 8 and 9 who score below proficiency can go back to seventh grade. Students who change schools or districts carry their baseline score with them.

A third-grade student who scored below proficiency has four years of allowable growth and must make up one-quarter of the gap between his or her third-grade score and the seventh-grade proficiency cut score in the first year, one-half of this proficiency gap by the end of the second year, three-fourths of the gap by the end of the third year, and must score at or above the cut score in the fourth year in order to be counted as proficient for AYP determinations. Similarly, a seventh-grader scoring below proficiency must close the gap by one-third in grade 8 and by two-thirds by the end of grade 9.[21] A student's baseline score is not reset in terms of the prior year score as long as the student continues to score below the proficiency cut score. However, the student must score at least as high as he or she did in the prior year as well as achieve the requisite gap reductions calculated from the baseline score in order to be counted as on-track to proficiency.

The Alaska growth model uses a "status-plus-growth" method for determining whether reporting groups meet the AMOs for language arts and mathematics. That is, first, the reporting group is assessed in terms of the percentage of students who scored at or above proficiency. If that percentage met the AMO, the group was classified as "met by status." If the percentage did not meet the AMO but the percentage not proficient was 10 percent or more less than the percentage not proficient in the prior year, the group was classified as "met by safe-harbor." If the group did

[21] Note that a ninth-grader must close the proficiency gap by 100 percent by the tenth grade, meaning that the growth model is equivalent to the status model for all ninth-graders who score below proficiency.

not meet the AMO by status or pass by safe-harbor, then the number of below-proficient students who are on-track to proficiency was added to the number of proficient students and this sum was divided by the total number of students in the reporting group who took the test. If this percentage met the AMO, the group was classified as "met by growth." If one or more reporting groups met by growth and the other AYP criteria noted below were also met, the school as a whole was classified as "made AYP by growth."

Alaska's *ESEA* accountability rules require a reporting subgroup to have more than 25 students before it is included in school AYP determinations. Subgroups larger than 40 students count only if 95 percent of them participate in testing, while smaller subgroups require that no more than two students fail to participate. Subgroups that meet the AMO within a 99 percent confidence interval, which varies by subgroup size, are determined to have made AYP.

The AYP results for Alaska in the 2006–07 and 2007–08 school years are shown in Exhibit 10. Of the 292 schools that met AYP requirements (222 by status plus 70 by safe-harbor) for that year, representing 59 percent of all eligible schools, no school made AYP via the state's pilot growth model. In 2006–07, 323 schools met AYP requirements, representing 66 percent of all eligible schools, and again no school made AYP via the state's pilot growth model.

Exhibit 10
Alaska School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07 and 2007–08

School AYP Under Status-Plus-Growth Model	Number		Percent	
	2006–07	2007–08	2006–07	2007–08
Met with Status	323	222	66%	45%
Met with Safe-Harbor	0	70	0%	14%
Met with Growth	0	0	0%	0%
Not Met	169	203	34%	41%
All Eligible Schools	492	495	100%	100%

Exhibit reads: For Alaska's schools overall, 222 met AYP under status in 2007–08, which was 45 percent of all eligible schools.

Source: *U.S. Department of Education, ED*Facts.

Arizona's Growth Model

Arizona's participation in GMPP was approved on July 3, 2007, for use beginning in the 2006–07 school year. The Arizona growth model includes students in grades 4 through 7 and uses scores on the Arizona Instrument to Measure Standards (AIMS) tests for reading and mathematics, which are vertically scaled for grades 3 through 8. Proficiency cut scores are set for each grade, and a student scoring below the cut score for either reading or math has three years or by eighth grade, whichever comes first, to reach the cut score.

The state uses a trajectory model that sets growth targets by dividing the difference between initial score and proficiency cut score three grades later into equal parts. In order to be counted as "on-track" for AYP purposes, a student in grades 3 through 5 must make up one-third of the shortfall in the first year, two-thirds of the shortfall by the end of the second year, and be proficient by the end of the third year. A sixth-grader must cover half of the shortfall in each of the two years of eligibility remaining.[22] Students who leave the Arizona school system before reaching proficiency have new growth targets set upon their return.

For the purposes of its growth model, Arizona calculates predicted AIMS scores by regressing current year's scores on previous year's scores and unique identifiers of the schools in which the students are enrolled. The student's actual previous year score and school identifier are multiplied by the respective regression coefficients to calculate a predicted score. These predicted scores are then subject to a 95 percent confidence interval, and the lower bound of that confidence interval is used as the student's score for comparing to his or her growth target. This regression-based adjustment procedure is done in order to correct for improvement that might be due to chance. This adjusting process has the effect of discounting high gains among lower-performing students and discounting declines among higher-performing students, because those patterns are anomalous with respect to the regression equation.

The state also uses a 99 percent confidence interval around annual AMOs for AYP determinations under the status model.[23] *ESEA* subgroups are counted in such determinations only if they have at least 40 students and if no fewer than 95 percent of these students participate in annual testing. Students who have estimated scores above the proficiency cutoff but who also fail to meet their growth targets continue to be counted as proficient for AYP purposes.

Arizona applied the results of the growth model after applying status and safe-harbor criteria. Growth model results were only applied to *ESEA* reporting groups that did not make AYP by status or safe-harbor. For those groups the number of students identified as "on-track to proficiency" was added to the number of proficient students and that sum was divided by the number of test-takers in the reporting group. That percentage was then compared to the AMO. If the school met all AYP criteria and had one or more reporting groups meeting the AMO because of the addition of on-track students, the school as a whole was classified in ED*Facts* as making AYP by growth.

ED*Facts* data for the 2007–08 school year indicate that Arizona's growth model resulted in only eight schools making AYP that would have missed if only status and safe-harbor were applied (Exhibit 11). Of the other 1,117 schools that met AYP requirements, 97 made AYP using the safe-harbor criteria, and 1,020 schools did so using status criteria alone. Data for the 2006–07 school year indicate that just one school made AYP via the growth model, while no schools made AYP via safe-harbor.

---

[22] Note that a seventh-grader must close the proficiency gap by 100 percent in the first year, meaning that the growth model is equivalent to the status model in this grade.

[23] Confidence intervals around the AMO (as opposed to around the percent proficient) are calculated using the number of full academic year students in the reporting group as the basis for the estimate. If the percent proficient in the group meets or exceeds the lower bound of this confidence interval, the group is classified as meeting the AMO.

Exhibit 11
Arizona School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07 and 2007–08

School AYP Under Status-Plus-Growth Model	Number		Percent	
	2006–07	2007–08	2006–07	2007–08
Met with Status	1,096	1,020	77%	68%
Met with Safe-harbor	0	97	0%	6%
Met with Growth	1	8	<1%	1%
Not Met	333	371	23%	25%
All Eligible Schools	1,430	1,496	100%	100%

Exhibit reads: For Arizona's schools overall in 2007–08, 1,020 met AYP under status, which was 68 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

Arkansas' Growth Model

Arkansas received approval to implement its proposed growth model in the 2006–07 school year. Arkansas uses a trajectory model that calculates growth for students in grades 4 through 7 using results of the Arkansas Benchmark Exams for mathematics and literacy, which are administered in grades 3 through 8 (plus grade 11 for literacy). Proficiency levels for these vertically scaled exams are set for each grade, and growth targets are based on the annual exam score increment needed to reach the proficiency standard in eighth grade. The annual increment that a student must attain in order to be classified as on-track to proficiency is calculated using grade-specific growth target multipliers of 0.295 in fourth grade, 0.319 in fifth grade, 0.385 in sixth grade, and 0.542 in seventh grade. These multipliers indicate the proportion of the total difference between the eighth-grade standard (a score of 700) and the student's current score that the student must gain over the next year in order to be on-track for eighth-grade proficiency. They contrast with the multipliers in other trajectory states (which all represent fractions of whole years—e.g., one-fourth, one-third, one-half) because Arkansas Benchmark Exams are not scaled to have a linear progression from one year's cut point to the next (see Appendix B for more detail and illustrations).

An important feature of this model is that it resets the growth target every year rather than setting a series of annual targets based on the first below-proficient exam score. For example, a third-grader who scores a 480 on the Arkansas Benchmark Exams is 20 points below the standard for that grade. Since the eighth-grade standard is 700, the student must gain 65 points [= (700–480) * 0.295] to be counted as on-track to proficiency. If the student scores a 539 in fourth grade (again 20 points below the standard), he or she fails to reach this threshold and now needs to gain 51 points [=(700–539) * 0.319] in fifth grade to be counted as on-track to proficiency. Arkansas students who make sufficient growth are counted as proficient for AYP determinations.

Arkansas applies confidence intervals to the grade-specific AMOs. Most of the other states in the GMPP apply confidence intervals to the percentages of proficient students rather than to the

AMO. The confidence interval is 95 percent and is applied by using the number of students enrolled in the school's tested grades for each reporting subgroup.

The data reported to ED*Facts* by Arkansas for the 2007–08 school year show that 15 percent of the schools made AYP according to the status model, 41 percent made AYP under safe-harbor provisions, and 6 percent made AYP under the growth model (Exhibit 12). These 6 percent represent schools that would not have made AYP had the GMPP not been available. In 2006–07, 8 percent of Arkansas schools made AYP via growth, while 24 percent made AYP under the status model and 32 percent made AYP using safe-harbor provisions.

Exhibit 12
Arkansas School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07 and 2007–08

School AYP Under Status-Plus-Growth Model	Number		Percent	
	2006–07	2007–08	2006–07	2007–08
Met with Status	212	131	24%	15%
Met with Safe-Harbor	287	369	32%	41%
Met with Growth	69	52	8%	6%
Not Met	326	338	36%	38%
All Eligible Schools	894	890	100%	100%

Exhibit reads: For Arkansas' schools overall, 131 met AYP under status in 2007–08, which was 15 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

The percentages of Arkansas schools making AYP with safe-harbor were much higher than in any of the other GMPP states in both years. It is not clear why this occurred.

Delaware's Growth Model

Delaware also received approval to use its proposed growth model starting with the 2006–07 school year. Delaware's model includes students in grades 3 through 10 and does not limit the number of years students have to make proficiency. The Delaware Student Testing Program (DSTP) yields vertically aligned scale scores for mathematics and language arts (a combination of reading and writing). Mathematics and reading tests are administered starting in grade 2 and growth can be assessed in grade 3 and above, but the writing component of the language arts scores does not start until grade 3.

The Delaware model is an example of the transition matrix type. Delaware uses a "value table" method for AYP determinations that assigns points for students depending on the type and extent of changes between the performance levels (see Exhibit 6). Points in the value table increase with the level of proficiency, so that a student scoring at the bottom receives fewer points for moving up one level (150) than a student moving from that level to the next (175). Conversely, students who move up two levels are awarded more points (225) than a student who moves to the same level in one step (175 or 200). All students who surpass their grade level proficiency cut score receive 300 points regardless of how high they score or whether their underlying scores actually declined from the prior year. AYP is determined by calculating the average points for the school and its subgroups in the two required subject areas and comparing these averages with the state's AMO levels.

Delaware was unique in the pilot project in that it combined results of the status and growth models, and classified all schools that made AYP using the combined results as "met AYP by growth" even if they would have met AYP by status alone. The other states employed the status-plus-growth model shown in Exhibit 4. Delaware's arrangement made it impossible to determine from ED*Facts* whether the schools listed as making AYP under growth also made AYP under status.

Drawing on the state reports in ED*Facts*, 87 of 183 schools in Delaware made AYP under the growth model in the 2007–08 school year (Exhibit 13). As a result of using the procedure of employing the growth model results before applying the status and safe-harbor provisions, the percentage of Delaware schools reported in ED*Facts* as making AYP under growth (48 percent) is much higher than most of the other pilot states. Many of these 87 schools would have made AYP under status or safe-harbor, had those criteria been applied before (as was done in the other GMPP states) the growth criteria. In 2006–07, a similar 46 percent of the schools made AYP under growth before applying status and safe-harbor.

Exhibit 13
Delaware School AYP Determinations With Growth Model Results Augmented With Status and Safe-Harbor Results, 2006–07 and 2007–08

School AYP Under Status-Plus-Growth Model	Number		Percent	
	2006–07	2007–08	2006–07	2007–08
Met with Status	21	41	12%	22%
Met with Safe-Harbor	19	0	11%	0
Met with Growth	83	87	46%	48%
Not Met	57	55	31%	30%
All Eligible Schools	180	183	100%	100%

Exhibit reads: For Delaware's schools overall, 41 met AYP under status in 2007–08, which was 22 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

The ED*Facts* reporting on school AYP outcomes used mutually exclusive categories of "by status" and "by growth" and it was thus not possible to use those data to address the question of how many Delaware schools that made AYP under their growth model would have missed AYP under the status model or by safe-harbor. However, data provided by the state directly to the evaluation team indicate that only five schools, or 3 percent of the schools in the state, made AYP because of the growth model provisions in both the 2007–08 and 2006–07 school years

(Exhibit 14).[24] These were the only schools that were reclassified from not making AYP to making AYP as a direct result of Delaware's participation in the GMPP.

Exhibit 14
Delaware School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07 and 2007–08

School AYP under Status-plus-Growth Model	Number		Percent	
	2006–07	2007–08	2006–07	2007–08
Met with Status	101	123	55%	67%
Met with Safe-harbor	22	0	12%	0%
Met with Growth	5	5	3%	3%
Not Met	57	55	31%	30%
All Eligible Schools	185	183	100%	100%

Exhibit reads: For Delaware's schools overall, 123 met AYP under status in 2007–08, which was 67 percent of all eligible schools.

Source: Delaware Department of Education.

Florida's Growth Model

Florida's growth model proposal was approved for use starting with the 2006–07 school year. The growth model applies to students in grades 3 through 10 using the Developmental Scale Scores (DSS) from the Florida Comprehensive Assessment Tests (FCAT) for mathematics and reading. The state uses a trajectory model that bases growth targets on the score required for proficiency three years after the first year tested (normally, grade 3). To be counted as proficient, a student who was not proficient at the baseline must close the gap between his or her baseline score and the proficiency cut score three grades later by one-third the first year and two-thirds the second year.[25] Students who continue to score below the cut score after three years start the process over and have new growth targets set.

Florida retains third-graders who score at or below level 1 (the lowest level) on the reading portion of the FCAT and who do not qualify for an exemption from this policy. The growth model incorporates such students by using their two third-grade scores on the FCAT DSS. Otherwise, only students in the fourth grade or later will have the two scores required to calculate growth.

For determining school AYP, Florida applies growth-model results for students only in *ESEA* reporting groups that do not meet their AMOs by status or safe-harbor. An important feature of Florida's accountability model is that, for groups not meeting their AMO, the growth-model data

[24] These data were provided to NORC by the Delaware Department of Education's Assessment and Accountability Branch. The data differed from the ED*Facts* reports in that both status and growth results were included for all schools and *ESEA* reporting groups.

[25] This means that a student can only make proficiency by growth for two years, because the gap must be fully closed in the third year; i.e., a student must meet or exceed the minimum proficiency score in year 3. A student first enrolled in grade 9 must close the gap by half to be counted as "on-track."

are used exclusively. That is, the number of non-proficient but on-track students is not added to the number of proficient students as is done in several other pilot states. Instead, the growth model results for both the proficient and non-proficient students are used. This means that students who are proficient but who are not on-track to maintain proficiency in three years are counted as "not proficient" for AYP purposes in these reporting groups. Groups not meeting AMO by status or safe-harbor must thus meet the AMO entirely on the basis of "on-track" students.

The AYP results for Florida in 2007–08 are shown in Exhibit 15. A total of 24 percent of Florida schools made AYP, with 16 percent making it under status, 4 percent by safe-harbor, and 5 percent by growth. The previous year, 34 percent of Florida schools made AYP, including 5 percent that made AYP via the pilot growth model. The overall percentage of schools that made AYP in Florida is much lower than the other GMPP states, but this may reflect higher standards for proficiency or higher AMOs rather than lower performance.

Exhibit 15
Florida School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07 and 2007–08

School AYP Under Status-Plus-Growth Model	Number		Percent	
	2006–07	2007–08	2006–07	2007–08
Met with Status	770	516	24%	16%
Met with Safe-harbor	155	116	5%	4%
Met with Growth	149	153	5%	5%
Not Met	2,135	2,495	67%	76%
All Eligible Schools	3,209	3,280	100%	100%

Exhibit reads: For Florida's schools overall, 516 met AYP under status in 2007–08, which was 16 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

Iowa's Growth Model

Iowa's growth model was approved for use with students in grades 3 through 8 beginning in the 2006–07 school year. Iowa uses third-grade math and reading scores on the Iowa Test of Basic Skills (ITBS) as a baseline, so growth calculations begin in fourth grade. Iowa's growth model is a type of transition matrix model (see Exhibit 5). To calculate growth, Iowa uses two categories of proficient (Intermediate and High) and three categories of below proficient (Weak, Low Marginal, and High Marginal). Category boundaries are set using national percentile ranks for each grade, with ITBS scale scores in the 40th percentile considered below proficient and 10th percentile scores considered Weak. Non-proficient students can still make Adequate Yearly Growth (AYG) if they improve to a higher category of below proficient within four years of the first year tested.

An important feature of Iowa's model is that students cannot fall back to a non-proficient category and still make AYG. This means that High Marginal students who decline to Low Marginal or Weak cannot make AYG simply by regaining High Marginal status, and that Low

Marginal students who decline to Weak must score in the High Marginal category to make adequate growth. Iowa counts all students making AYG as proficient for *ESEA* reporting groups that fail to make AYP when the status and safe-harbor provisions have been applied.

The 2007–08 AYP results for Iowa show that 2 percent of the schools made AYP with the growth model in the status-plus-growth framework (Exhibit 16). This was on top of the 62 percent of schools making AYP under status and 4 percent via safe-harbor. In 2006–07, 11 percent of Iowa schools made AYP using growth criteria, 75 percent made AYP under status, and 10 percent made AYP via safe-harbor.

Exhibit 16
Iowa School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07 and 2007–08

School AYP Under Status-Plus-Growth Model	Number		Percent	
	2006–07	2007–08	2006–07	2007–08
Met with Status	831	678	75%	62%
Met with Safe-harbor	105	43	10%	4%
Met with Growth	116	23	11%	2%
Not Met	52	354	5%	32%
All Eligible Schools	1,104	1,098	100%	100%

Exhibit reads: For Iowa's schools overall, 678 met AYP under status in 2007–08, which was 62 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

North Carolina's Growth Model

North Carolina received approval on May 17, 2006, to use its growth model for the 2005–06 school year. The state uses a trajectory model that calculates growth in grades 3 through 7 using vertically equated North Carolina end-of-grade tests for mathematics and reading. Third-graders in North Carolina have four years to grow to proficiency, because they take a test upon entering the third grade and a test at the end of the year. All other students take end-of-grade tests and thus have only three years or until eighth grade to become proficient, whichever comes first.

For students who take the third-grade pretest, growth targets are set by dividing the difference between the initial test score and the proficiency cut score for sixth grade into four equal parts. Thus third-grade students who make up one-fourth of the shortfall between the baseline score and sixth-grade standards by the end of third grade are considered to be "on-track" to proficiency. Students who enter the school system after third grade must close this gap more quickly, because they will have a three-year horizon before eighth grade, when all students are expected to make a proficient score.

For AYP purposes, North Carolina applies growth-model results after status and safe-harbor criteria. For reporting groups not meeting their AMO by status or safe-harbor, the number of non-proficient students who are on-track to proficiency per the growth model is added to the number of proficient students in the group. A group that reaches the state's AMO with the inclusion of "on-track" students is considered to have met the AMO by growth and the school as a whole is classified as having made AYP by growth.

North Carolina applies a 95 percent confidence interval before using the growth model for AYP determinations and only includes subgroups with at least 40 full academic year (starting on Oct. 1 of the current school year) students. The standard 95 percent participation rule also applies to subgroups and schools and is based on the full set of students enrolled in the current school year.

Exhibit 17 shows that the pilot growth model resulted in no more North Carolina schools making AYP in 2007–08 than would have made AYP through status (399) and safe-harbor (338) alone. More than two-thirds (69 percent) of schools in North Carolina did not meet AYP requirements, up from just over half (56 percent) of schools in 2006–07. Twelve schools also met their AYP requirements via the pilot growth model during that school year.

Exhibit 17
North Carolina School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07 and 2007–08

School AYP Under Status-Plus-Growth Model	Number		Percent	
	2006–07	2007–08	2006–07	2007–08
Met with Status	661	399	30%	17%
Met with Safe-harbor	308	338	14%	14%
Met with Growth	12	0	1%	0%
Not Met	1,226	1,612	56%	69%
All Eligible Schools	2,207	2,349	100%	100%

Exhibit reads: For North Carolina's schools overall, 399 met AYP under status in 2007–08, which was 17 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

Ohio's Growth Model

Ohio was notified of its admission into the pilot program on August 15, 2007, subject to the condition that the state adopt a uniform minimum subgroup size. Such a resolution passed the Ohio General Assembly on April 23, 2008. Ohio's approved plan uses a projection model for determining whether students are on-track to proficiency. It includes growth calculations for grades 3 through 7 using scaled scores for reading and mathematics on the Ohio Achievement Tests plus grade 10 scores on the Ohio Graduation Tests.[26] Performance standards for each subject and grade are expressed as cut score points. Students have three years to make a proficient cut score, fewer if the student is expected to graduate from the current school before then. A student is considered to be on-track to proficiency if he or she is "projected" to meet the

[26] Ohio used (equated) 2004–05 data from the Ohio Proficiency Tests for eighth-graders in 2006–07 to give them at least three years of test results.

cut score in three years or for the next grade beyond the current school's terminal grade, whichever comes first.

These projections of student achievement are derived from a complex statistical model designed to predict each student's level based on data about his or her past performance and the school the student will most likely attend after his or her current school. The statistical model uses at least three and up to five years of all available test scores. The model uses regression analysis to calculate the relationships between the past scores and endpoint school and the target endpoint score for a "reference cohort" defined as the most recent cohort of students that completed the endpoint. Consistent with the GMPP core principles (see Exhibit 3), the model does not make any adjustments for demographic characteristics, meaning that students with differing socioeconomic or racial backgrounds but with the same prior academic achievement and likely next school will have the same projections.

The regression coefficients estimated with the reference cohort are then used to calculate predicted endpoint scores for the students in the younger cohorts. Standard errors around the predictions are also estimated and two standard error units are added to each predicted score.[27] This adjustment reduces the chance of misclassifying a student who is actually on-track as not on-track, but increases the chance of misclassifying a student who is actually not on-track as on-track. If the adjusted prediction meets or exceeds the cut score for the target grade, the student is classified as on-track. An important feature of Ohio's growth model is that projections are possible even if a student is missing prior achievement scores for some grades or subjects. Also, the model recalculates projections every year and calculates projections for students who score above proficiency.

Ohio uses the growth model results for AYP determinations only after applying status and safe-harbor criteria. For subgroups that do not make AYP by status or safe-harbor, students scoring below proficiency but who are identified as on-track are counted the same as proficient students. If the proportion of proficient plus on-track students meets or exceeds the AMO, then the subgroup is classified as "making AYP by growth" in the federal reporting system. If all subgroups make AYP by an available method and at least one subgroup makes AYP by growth, then the school as a whole is classified as making AYP by growth in the ED*Facts* system.

Ohio joined the GMPP in the 2007–08 school year and thus has only one year of results. Exhibit 18 shows that the pilot growth model resulted in 983 more Ohio schools making AYP in 2007–08 than would have made AYP through status (798) and safe-harbor (163) alone. About a third (34 percent) of schools in Ohio did not meet their AYP requirements by status, safe-harbor, or growth. Ohio had by far the largest reported percentage of schools making AYP through the growth model (34 percent) of all the pilot states; the states with the next highest percentages were Arkansas (6 percent) and Florida (5 percent).

[27] Ohio Department of Education and SAS Institute, Inc., Conference Call, July 19, 2010.

Exhibit 18
Ohio School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2007–08

School AYP Under Status-Plus-Growth Model	Number	Percent
Met with Status	798	27%
Met with Safe-harbor	163	6%
Met with Growth	983	34%
Not Met	984	34%
All Eligible Schools	2,928	100%

Exhibit reads: For Ohio's schools overall, 798 met AYP under status, which was 27 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

Tennessee's Growth Model

Tennessee received approval on May 17, 2006, to use its growth model beginning in the 2005–06 school year.[28] The state calculates growth for students in grades 4 through 8 using vertically aligned scores for mathematics and reading or language arts from the Tennessee Comprehensive Assessment Program (TCAP). Students are first tested in the third grade, and those who score below proficient are given until ninth grade to reach proficiency by growth. Tennessee's projection model establishes growth targets for students by calculating a predicted ninth-grade achievement score for each student on the basis of his or her current and prior test scores. As in Ohio, predicted scores are calculated on the basis of regression coefficients estimated using data from a standard-setting cohort that has just completed the eighth grade. Unlike Ohio, however, Tennessee does not adjust the predicted scores with estimates of standard errors. Students who score below proficiency but who are predicted to be above the TCAP cut scores within three years or by grade 9, whichever comes first, are counted as proficient for AYP growth determinations.

For students past the fifth grade, proficiency cut scores are based on the TCAP high school graduation (or Gateway) assessments. As is the case with the Ohio growth model, the Tennessee growth model is able to make projections even if a student is missing prior achievement scores for some grades or subjects. The Tennessee model also recalculates projections every year and calculates projections for students who score above proficiency.

For AYP purposes, the growth-model results are applied only to *ESEA* reporting groups that did not meet their AMOs by status or safe-harbor. As in Florida, Tennessee uses growth-model results for all students in groups not meeting their AMO. That is, proficient students in those groups who are projected to score below the cutoff in three years are not counted as proficient. A school in Tennessee making AYP is considered to have made it by growth if including

[28] This approval stipulated that Tennessee work with the U.S. Department of Education to ensure that results from alternate assessments using alternate achievement standards were included by the 2006–07 school year.

students projected to be proficient allows all non-proficient subgroups to meet the AMO. Tennessee includes subgroups in AYP determinations if they have at least 45 full academic year students and applies a 95 percent confidence interval to the AMO for status and safe-harbor determinations (like Arizona and Arkansas) instead of to the percent proficient.

Exhibit 19 shows that 1,130 (or 83 percent) of 1,357 Tennessee schools met AYP requirements in 2007–08 through status and safe-harbor criteria. An additional 22 schools (or 2 percent) that failed to meet these criteria made AYP by also applying the pilot growth model. A similar number of schools (19) made AYP by growth the previous year, though the total number of schools making AYP was higher (90 percent).

Exhibit 19
Tennessee School AYP Determinations With Status and Safe-Harbor Results Augmented With Growth Model Results, 2006–07 and 2007–08

School AYP Under Status-Plus-Growth Model	Number		Percent	
	2006–07	2007–08	2006–07	2007–08
Met with Status	1,149	964	84%	71%
Met with Safe-harbor	68	166	5%	12%
Met with Growth	19	22	1%	2%
Not Met	126	205	9%	15%
All Eligible Schools	1,362	1,357	100%	100%

Exhibit reads: For Tennessee's schools overall, 964 met AYP under status in 2007–08, which was 71 percent of all eligible schools.

Source: U.S. Department of Education, ED*Facts*.

Impact of GMPP on AYP

As can be seen in Exhibits 10–19, the pilot states varied in the proportion of their schools making AYP under status or safe-harbor; these differences can obscure the impact of their growth models on AYP. One way to assess the impact is to calculate the percentage increase in the number of schools making AYP due to use of the growth model. The schools making AYP uniquely by growth represented a percentage increase in the schools making AYP of 20 percent across all states, and ranged as high as 102 percent in Ohio, 24 percent in Florida, and 10 percent in Arkansas (Exhibit 20). Excluding Ohio, the overall percentage increase due to the GMPP in the other eight states was 5 percent.

Another indicator of the impact of the GMPP is the extent to which the growth models affected the pool of schools identified as failing both the status and safe-harbor criteria for AYP. Across all nine states, 16 percent of the 7,863 schools that did not make AYP by status or safe-harbor made AYP by growth (Exhibit 20). These rates ranged from a high of 50 percent of eligible schools in Ohio making AYP by growth to 5 percent or fewer of eligible schools in Alaska, Arizona, and North Carolina making AYP by growth. Excluding Ohio, the overall percentage reduction due to the GMPP in the other eight states was just 4 percent compared to the 16 percent when Ohio was included. As noted in the section above focusing on features of Ohio's growth model, the main reason for the much higher impact of the growth model there on school

AYP was that the state used a much more inclusive definition of on-track to proficiency than the other pilot states.

Exhibit 20
Percentage Increase in Number of Schools That Made AYP Due to Growth, and Percentage Decrease in Number of Schools That Did Not Make AYP Due to Growth, by State, 2007–08

Pilot States	Number of Schools Making AYP by Status or Safe-Harbor	Number of Schools not Making AYP by Status or Safe-Harbor	Number of Schools Not Making AYP by Status or Safe-Harbor That Met by Growth	Percentage Increase in Schools Making AYP Due to Growth	Percent of Schools Not Making AYP by Status or Safe-Harbor That Met by Growth
All Nine States	6,213	7,863	1,246	20%	16%
Alaska	292	203	0	0%	0%
Arizona	1,117	379	8	1%	2%
Arkansas	500	390	52	10%	13%
Delaware	123	60	5	4%	8%
Florida	632	2,648	153	24%	6%
Iowa	721	377	23	3%	6%
North Carolina	737	1,612	0	0%	0%
Ohio	961	1,967	983	102%	50%
Tennessee	1,130	227	22	2%	10%

Exhibit reads: The 1,246 schools that made AYP by growth increased the number of schools making AYP from 6,213 to 7,459 schools, which was a percentage increase of 20 percent. Of the 7,863 schools that did not make AYP under either status or safe-harbor, 1,246 or 16 percent made AYP using the growth model.

Source: U.S. Department of Education, ED*Facts* and the Delaware state department of education.

Impact of the GMPP on Subgroup AYP Outcomes

ESEA requires that each of a number of subgroups must reach AMOs in order for a school to receive AYP credit. This section documents the extent to which the growth models affect the AYP outcomes of each targeted subgroup. For each group, we present the numbers and percentages of subgroups that did not reach the AMO by the status model or safe-harbor that reached the AMO because of growth model.

Although schools do not in most cases miss AYP simply because of the performance of a single subgroup,[29] growth models may be especially beneficial to historically underperforming subgroups. This analysis provides a more fine-grained picture of the impact of the growth models in each state, pinpointing the rates at which each subgroup reached the AMO because of the growth model.

[29] See U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service (2007). *State and Local Implementation of the No Child Left Behind Act, Volume III—Accountability Under NCLB: Interim Report*: 42, 43.

The school-level AYP determinations and the methods of making AYP (status, safe-harbor, growth) examined thus far are based on results for the complete set of *ESEA* reporting subgroups represented in each school. The states also reported to ED*Facts* whether each subgroup was exempt by minimum group size criteria, met the AMO by status, met the AMO by safe-harbor, met the AMO by growth, or did not meet the AMO for both reading and mathematics. A subgroup met the AMO by growth if it did not make AYP by status or safe-harbor and sufficient numbers of below-proficient students were on-track to reach the AMO.

The number and percentage of each subgroup that was classified as meeting the AMO in reading by growth are shown in Exhibit 21 (results for mathematics are generally consistent with the reading results and are presented in Appendix C). In this analysis, schools are considered to be eligible to meet their AMO by growth if the respective reporting subgroup did not meet the AMO by either status or safe-harbor. The numbers of subgroups that met the AMO by growth are greater than the number of schools that made AYP by growth because schools may have one or more subgroups that met their AMO by growth but still had at least one other subgroup that did not meet the AMO by any method. If any subgroup within a school failed to meet the AMO by any of the three methods, the school would not make AYP.

Exhibit 21
Number of Eligible Schools and Percentage in Which Reading AMO Was Met Because of Growth Model Results, by All Students and Racial or Ethnic Reporting Groups, 2007–08

	Number of Eligible Schools*					
Pilot States	**All Students**	**White**	**Black, Non-Hispanic**	**Hispanic**	**Asian/Pacific Islander**	**American Indian/ Alaskan Native**
All Nine States	1,818	373	1,927	737	47	181
Alaska	74	3	3	0	8	68
Arizona	4	37	58	46	37	99
Arkansas	48	9	75	4	0	0
Delaware**	8	NA	NA	NA	NA	NA
Florida	849	108	922	484	1	0
Iowa	71	27	41	29	1	3
North Carolina	196	2	316	103	0	11
Ohio	529	171	487	69	0	0
Tennessee	39	16	25	2	0	0
	Percent of Eligible Schools					
All Nine States	18%	39%	15%	12%	0%	0%
Alaska	0%	0%	0%	NA	0%	0%
Arizona	0%	0%	0%	0%	0%	0%
Arkansas	0%	0%	0%	0%	NA	NA
Delaware	0%	NA	NA	NA	NA	NA
Florida	6%	<1%	6%	9%	0%	NA
Iowa	11%	11%	10%	7%	0%	0%
North Carolina	<1%	0%	<1%	0%	NA	0%
Ohio	48%	74%	42%	61%	NA	NA
Tennessee	51%	88%	52%	0%	NA	NA

Exhibit reads: Across all nine states, among the 1,818 schools in which the 'all students' reporting group did not reach the reading AMO by either status or safe-harbor, that group did reach the AMO in 18 percent of the eligible schools when the growth criteria were applied.

* "Eligible schools" means schools in which the respective reporting group did not meet the reading AMO by status or safe-harbor and had grade levels to which the growth model was applied.

** ED*Facts* did not include subgroup information for Delaware in 2007–08.

NA: Not applicable due to no eligible schools.

Source: U.S. Department of Education, ED*Facts*.

Exhibit 21 shows that subgroups meeting AMOs by growth were concentrated in Florida, Iowa, Ohio, and Tennessee, with the highest rates in Ohio and Tennessee. Overall, more black (1,927) and Hispanic subgroups (737) did not meet their reading AMO by status or safe-harbor than white subgroups (373), but lower percentages of the eligible black and Hispanic subgroups met the AMO when growth model results were factored in (15 percent of black and 12 percent of Hispanic subgroups versus 39 percent of the eligible white subgroups). Florida is an exception to this trend, with relatively fewer white subgroups (less than 1 percent) in eligible schools benefiting from the state's pilot growth model than black (6 percent) and Hispanic subgroups (9 percent). Similarly for mathematics, relatively more eligible Hispanic (15 percent) subgroups met the AMO by growth than white (10 percent) subgroups in Florida (see Exhibit C.2).

Exhibit 22 extends the picture to the remaining three *ESEA* reporting subgroups, showing the impact of the GMPP on economically disadvantaged students, students with disabilities (SWD), and limited English proficient (LEP) students. In all nine states combined, 21 percent of low-income subgroups that missed the reading AMO by status and safe-harbor were able to meet their AMO via the pilot growth models (results are similar for mathematics, see Exhibit C.3). Relatively fewer SWD (15 percent) and LEP student (6 percent) subgroups in eligible schools met the reading AMO by growth. The picture is slightly different in Florida and Iowa, where eligible LEP student subgroups met their AMO at higher rates from the GMPP than did low-income student subgroups.

The higher rates of subgroups meeting both the reading and mathematics AMOs by growth in Ohio and Tennessee than in the other pilot states are noteworthy in that both states used a projection model methodology for making on-track to proficiency determinations. However, two factors make the impact of the projection model per se somewhat unclear. First, the relative size of the eligible pools in those two states were very different because of the difference in cut scores; Tennessee had relatively low cut scores and high rates of schools making AYP by status or safe-harbor, while Ohio had relatively high cut scores and much lower rates of AYP by status or safe-harbor. Second, Ohio adjusted its projection estimated scores upward by two standard error units, while Tennessee did not make such an adjustment. If Ohio had not made that adjustment, it is likely it would have had a much lower rate of on-track students among its non-proficient students and thus lower rates of subgroups meeting AMOs by growth and schools making AYP by growth.

Exhibit 22
Number of Eligible Schools and Percentage in Which Reading AMO Was Met Because of Growth Model Results, by Low-SES, SWD, and LEP Reporting Groups, 2007–08

Pilot States	Number of Eligible Schools*		
	Economically Disadvantaged Students (low-SES)	Students with Disabilities (SWD)	Limited English Proficient (LEP) Students
All Nine States	2,562	3,017	833
Alaska	81	42	53
Arizona	71	282	157
Arkansas	76	70	6
Delaware**	NA	NA	NA
Florida	1,075	1,185	463
Iowa	143	142	27
North Carolina	363	402	83
Ohio	719	881	43
Tennessee	34	13	1
	Percent of Eligible Schools		
All Nine States	21%	15%	6%
Alaska	0%	0%	0%
Arizona	0%	0%	0%
Arkansas	0%	0%	0%
Delaware	NA	NA	NA
Florida	7%	4%	8%
Iowa	8%	15%	15%
North Carolina	<1%	0%	0%
Ohio	58%	42%	19%
Tennessee	50%	15%	0%

Exhibit reads: Across all nine states, among the 2,562 schools in which the "economically disadvantaged" reporting group did not reach the reading AMO by status or safe-harbor, that group did reach the AMO in 21 percent of the eligible schools when the growth criteria were applied.

* "Eligible schools" means schools in which the respective reporting group did not meet the reading AMO by status or safe-harbor and had grade levels to which the growth model was applied.

** ED*Facts* did not include subgroup information for Delaware in 2007–08.

NA: Not applicable due to no eligible schools.

Source: U.S. Department of Education, ED*Facts*.

Discussion

The idea of using growth models to assess school academic performance is attractive and the data collection systems and database architecture needed to support such models are rapidly becoming available. The pilot models were strictly add-ons to the traditional status-plus-safe-harbor models for determining AYP in all nine states. Even Delaware, the only state that applied the growth model and then augmented with the status model, used the growth model as an add-on instead of a replacement. While states had the option of applying growth criteria instead of status and safe-harbor, none exercised that option. Growth criteria were instead added to status and safe-harbor criteria in ways that could only increase the number of schools making AYP.

States also had the option of applying growth criteria before status and safe-harbor were applied, but Delaware was the only state that did so. Delaware's growth-plus-status procedure had the effect of officially recognizing about 48 percent of their schools as making AYP by growth, which was a much higher percentage than under the status-plus-growth procedure used in the other states except Ohio. While this may have had a positive effect on the public visibility and understanding of the model in Delaware, the actual number of Delaware schools labeled as making AYP by growth that would not have made AYP under status or safe-harbor was small.

As shown in Exhibits 10 through 19, the percent of schools that made AYP using the growth provisions of the GMPP, and which would not have made AYP without those provisions, ranged from a high of 34 percent of schools in Ohio to 0 percent in Alaska and North Carolina. Across all nine pilot states, most of the schools that made AYP by growth were located in Ohio; excluding Ohio, only 2 percent of all schools in the other eight states made AYP by growth.

Another perspective on these rates is gained when the number of schools making AYP under growth is compared not to the total number of schools in the state but instead to either the number of schools in the state that made AYP under either status or safe-harbor (the percentage increase due to growth) or the number that did not make AYP under either status or safe-harbor (percentage decrease in non-AYP due to growth). The percentage increases in schools making AYP ranged from a high of 102 percent in Ohio to less than 2 percent in Alaska, Arizona, North Carolina, and Tennessee. The growth model decreased the number of schools identified as failing under the status-plus-safe-harbor model by 50 percent in Ohio but by less than 5 percent in Alaska, Florida, and North Carolina. The rate of making AYP by growth after missing by status and safe-harbor ranged from 6 percent to 13 percent in the remaining five pilot states.

Reasons for the variation among states in the percentages of schools making AYP by growth will be examined systematically in Chapter IV but are generally a combination of differences in the assessments used, the proficiency cutpoints, the features of the growth models implemented, the states' AMOs, and the actual levels of growth realized by the students. It is noteworthy that the percentages making AYP by growth appear unrelated to the percentages making AYP by status or safe-harbor (both of which vary greatly among these states) or to whether the state adopted a transition matrix, trajectory, or projection type of growth model. The exceptionally high rate in Ohio is most likely explained by that state's practice of augmenting students' projected scores with statistically-derived quantities reflecting the uncertainty of the projected scores. In contrast, the very low rates in other states may reflect the impact of the various other (nongrowth) methods for determining AYP available in those states for schools to make AYP by status and safe-harbor (e.g., confidence intervals and multiyear averaging), such that the status and safe-harbor methods picked up schools which would have made AYP by growth had those various provisions not been available. This issue is addressed in Chapter IV, where the numbers of schools in each state that could make AYP if growth model results were given priority over status and safe-harbor are estimated.

III. School Characteristics Associated with AYP Outcomes

This chapter addresses the question of whether certain school characteristics are associated with making AYP under growth but not status or safe-harbor. The school characteristics are organizational and demographic variables that include *ESEA* improvement status, poverty level, minority concentration, locality, and size. Of particular interest here is how pilot growth models affect the AYP designation of schools serving disadvantaged populations. Because such schools have been found to make AYP at much lower rates than those serving more affluent populations,[30] the GMPP may function to lessen the AYP gap on this dimension.

The question addressed in this chapter is whether certain types of schools are more or less likely to make AYP by the growth criteria. To answer this question, the following analyses look at the percentage increase of schools of each type that made AYP because of the availability of the growth model under the GMPP. The percentage increase is calculated by dividing the number of schools making AYP by growth but not by status or safe-harbor by the number that made AYP by either status or safe-harbor. Any difference between comparison groups in the percentage increases of schools making AYP by growth will suggest that the GMPP had a disproportionate effect on certain types of schools.

The U.S. Department of Education report, *State and Local Implementation of the No Child Left Behind Act, Volume II—Accountability Under NCLB: Interim Report* (Le Floch, Martinez, O'Day, et al. 2007) found that certain school demographic characteristics were associated with the likelihood of making AYP. For example, the study found that high-poverty, high-minority, larger, and urban schools were less likely to make AYP by status criteria (Le Floch, Martinez, O'Day, et al. 2007: 39, 40). Another important characteristic considered here is the *ESEA* improvement status classifications of schools. This chapter uses the data reported in ED*Facts* to characterize schools that made AYP under growth but not under status or safe-harbor. As discussed in Chapter II, Delaware reported results to ED*Facts* based on growth model results augmented with status and safe-harbor results. In order to identify the incremental impact of the growth model on schools' likelihood of making AYP, we substitute data using status and safe-harbor augmented with growth for the Delaware ED*Facts* report.

ESEA *School Improvement Status.* An important issue for *ESEA* accountability is the extent to which schools identified for improvement are, despite their relatively low levels of performance, actually making progress toward the goal of universal proficiency. The growth model pilot is in part intended to identify such schools. States are required under *ESEA* to identify for improvement Title I schools that do not meet AYP for two consecutive years and to initiate a process of interventions designed to improve student outcomes; most states have applied a similar process for non–Title I schools. Schools not making AYP for two consecutive years are officially identified for improvement. If a school does not make AYP after two years in the improvement status, it is

[30] U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service (2007). *State and Local Implementation of the No Child Left Behind Act, Volume III—Accountability Under NCLB: Interim Report*: 39.

identified for corrective action. If the school misses AYP for an additional year, it moves into restructuring status.[31]

As is shown in Exhibit 23 below, states vary with respect to whether the use of growth models allowed relatively more schools identified for improvement to make AYP than schools not identified for improvement. The percentage increase in making AYP due to growth is greater for identified versus unidentified schools in three states (Arizona, Iowa, and Ohio), lesser in two states (Arkansas and Delaware), and only slightly different in two states (Florida and Tennessee). In North Carolina, no schools made AYP by growth regardless of status relative to being identified for improvement. Improvement status data for Alaska was not included in ED*Facts* for year 2007–08.

Exhibit 23
Numbers of Schools Making AYP by Status or Safe-Harbor, and Percentage Increase in Schools Making AYP Due to Growth, by *NCLB* School Improvement Status, 2007–08

	Numbers of Schools Making AYP by Status or Safe-harbor			Percentage Increase in Schools Making AYP Due to Growth		
Pilot States	Identified for Improvement/ Under Corrective Action	Planning to Restructure or Restructuring	Not Identified for Improvement	Identified for Improvement/ Under Corrective Action	Planning to Restructure or Restructuring	Not Identified for Improvement
All Eight States	359	56	5,466	69%	48%	18%
Arizona	45	4	1,068	7%	25%	<1%
Arkansas	76	10	414	5%	0%	12%
Delaware	3	0	120	0%	0%	4%
Florida	31	27	574	26%	48%	23%
Iowa	3	0	718	167%	∞*	2%
North Carolina	65	5	653	0%	0%	0%
Ohio	70	4	872	321%	300%	84%
Tennessee	66	6	1,047	3%	0%	2%

Exhibit reads: Of the schools across all eight pilot states that are either identified for improvement or under corrective action in 2007–08, 359 make AYP by status or safe-harbor and the number making AYP is increased 69 percent by schools making AYP by growth.

* One school in Iowa that was planning to restructure or in the process of restructuring made AYP by growth but none made AYP by status or safe-harbor. The percentage increase due to growth could not be calculated because the denominator was zero.

Source: U.S. Department of Education, ED*Facts*.

School Poverty Concentration. The 2007 Interim Report on the implementation of *NCLB* found that schools' likelihood of making AYP under the status model was strongly correlated with student socioeconomic status and racial or ethnic minority representation.[32] Exhibit 24 below shows that the

[31] U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service (2007). *State and Local Implementation of the No Child Left Behind Act, Volume III—Accountability Under NCLB: Interim Report*: 3-6.

[32] Ibid, 39.

GMPP reduced the gap between low- and high-poverty schools.[33] Expressed as a percentage increase, the pilot growth models increased the number of high-poverty schools making AYP by 20 percent while increasing the number of low-poverty schools making AYP by 18 percent. There is considerable variation by state, however. The effect of growth on high- compared with low-poverty schools making AYP was much higher in Florida (71 percent versus 10 percent), Iowa (20 percent versus 3 percent), Ohio (169 percent versus 68 percent) and Tennessee (6 percent versus <1 percent). Growth models had little or no differential impact on high-poverty schools making AYP in Alaska, Arizona, Arkansas, Delaware, and North Carolina.

Exhibit 24
Numbers of Schools Making AYP Under Status-Plus-Safe-Harbor, and Percentage Increase in AYP Due to Growth, by School Poverty Concentration, 2007–08

	Numbers of Schools Making AYP by Status or Safe-Harbor			Percentage Increase in Schools Making AYP Due to Growth		
Pilot States	**Low Poverty**	**Medium Poverty**	**High Poverty**	**Low Poverty**	**Medium Poverty**	**High Poverty**
All Nine States	1,946	3,402	826	18%	21%	20%
Alaska	276	0	16	0%	0%	0%
Arizona	315	520	281	<1%	<1%	1%
Arkansas	25	389	86	8%	11%	8%
Delaware	29	90	4	0%	6%	0%
Florida	272	315	45	10%	30%	71%
Iowa	265	451	5	3%	3%	20%
North Carolina	187	471	61	0%	0%	0%
Ohio	458	439	64	68%	128%	169%
Tennessee	119	727	264	<1%	<1%	6%

Exhibit reads: Of the schools across all nine pilot states that had low enrollments of children from poverty-level households in 2007–08, 1,946 make AYP by status or safe-harbor and the number making AYP is increased 18 percent by schools making AYP by growth.

Source: U.S. Department of Education, ED*Facts*.

School Minority Concentration. Results for school minority composition are shown in Exhibit 25.[34] Across all nine states the use of pilot growth models appears to benefit low-minority schools more than high-minority schools (23 percent versus 18 percent). Analyzed separately, however, all states show equal or greater growth impact for high-minority schools (an example of the “aggregation paradox”). The gap in making AYP between high- and low-minority schools was reduced in Arizona, Arkansas, Florida, Ohio, and Tennessee. In these states the percent increase for high-poverty schools making AYP by growth was considerably higher than the percent increase for low-minority schools. Two states—Delaware and Iowa—did not have any high-minority schools

[33] Poverty level is defined here in terms of the percentages of students eligible for the federal free and reduced-price lunch program (FRPL). Low-poverty schools enroll 25 percent or fewer FRPL-eligible students, medium-poverty schools enroll 26 percent to 75 percent FRPL-eligible students, and high-poverty schools enroll more than 75 percent.

[34] Minority level is defined here in terms of the percentages of non-white or Hispanic students. Low-minority schools enroll 25 percent or fewer non-white or Hispanic students, medium-minority schools enroll 26 percent to 75 percent non-white or Hispanic students, and high-minority schools enroll more than 75 percent.

but displayed a similar pattern in having the use of growth models close the AYP gap among medium- and low-minority schools. In Alaska and North Carolina, no schools made AYP by growth regardless of minority composition.

Exhibit 25
Numbers of Schools Making AYP Under Status-Plus-Safe-Harbor, and Percentage Increase in AYP Due to Growth, by School Minority Concentration, 2007–08

Pilot States	Numbers of Schools Making AYP by Status or Safe-Harbor			Percentage Increase in Schools Making AYP Due to Growth		
	Low Minority	Medium Minority	High Minority	Low Minority	Medium Minority	High Minority
All Nine States	3,894	1,553	645	23%	14%	18%
Alaska	123	78	86	0%	0%	0%
Arizona	368	499	213	<1%	<1%	2%
Arkansas	343	126	22	8%	15%	14%
Delaware	31	86	6	0%	6%	0%
Florida	275	236	110	11%	33%	37%
Iowa	684	21	0	3%	14%	0%
North Carolina	449	229	43	0%	0%	0%
Ohio	854	59	29	95%	180%	197%
Tennessee	767	219	136	<1%	4%	6%

Exhibit reads: Of the schools across all nine pilot states that had low enrollments of minority children in 2007–08, 3,894 make AYP by status or safe-harbor and the number of schools making AYP is increased 23 percent by schools making AYP by growth.

Source: U.S. Department of Education, ED*Facts*.

School Urbanicity. School location in an urban, suburban, or rural area is another variable that previous studies have found related to AYP results, with rural and suburban schools generally more likely to make AYP under status or safe-harbor than urban schools. Exhibit 26 shows that, in most states, the growth model moved relatively more urban schools than suburban or rural schools into making AYP. Iowa was the only state among the nine to exhibit the opposite trend, while the GMPP tended to benefit suburban (28 percent) versus urban (23 percent) schools in Florida. Again, Alaska and North Carolina had no schools making AYP by growth in any category.

Exhibit 26
Numbers of Schools Making AYP Under Status-Plus-Safe-Harbor, and Percentage Increase in AYP Due to Growth, by School Urbanicity, 2007–08

Pilot States	Numbers of Schools Making AYP by Status or Safe-Harbor			Percentage Increase in Schools Making AYP Due to Growth		
	Urban Schools	Suburban Schools	Rural Schools	Urban Schools	Suburban Schools	Rural Schools
All Nine States	1,327	1,341	3,460	16%	37%	15%
Alaska	34	6	248	0%	0%	0%
Arizona	471	213	399	1%	<1%	<1%
Arkansas	73	35	383	19%	14%	8%
Delaware	13	57	53	23%	2%	2%
Florida	152	327	146	23%	28%	16%
Iowa	75	43	597	1%	5%	3%
North Carolina	119	112	503	0%	0%	0%
Ohio	105	378	464	135%	103%	95%
Tennessee	285	170	667	3%	1%	2%

Exhibit reads: Of the schools across all nine pilot states that were located in urban areas in 2007–08, 1,327 make AYP under status or safe-harbor and the number of schools making AYP is increased 16 percent by schools making AYP by growth.

Source: U.S. Department of Education, ED*Facts*.

Discussion

The answer to the question, "Is the likelihood of making AYP under growth models associated with school characteristics?" varied across the pilot states and also varied within states across years. Exhibit 27 summarizes the results in Exhibits 23 through 26 above by showing whether the GMPP reduced the AYP gap (√) for that type of school compared with schools at the other end of the spectrum (e.g., high- vs. low-poverty schools). Results for 2006–07 published in the *Interim Report on the Evaluation of the Growth Model Pilot Project* (2010) are also added to compare the impact of the pilot program across years. The results summarized here are all bivariate relationships and do not indicate independent effects of the school characteristics based on statistical adjustments for the substantial overlaps among the characteristics (for example, the tendency of schools identified for improvement to also enroll high proportions of low-income minority students).

Exhibit 27
Summary of the Effect of GMPP by School Demographic Characteristic and by State, 2006–07 and 2007–08

Pilot States	Identified for Improvement / Under Corrective Action		High Poverty School		High Minority School		Urban School (compared to suburban schools)	
	2006–07	2007–08	2006–07	2007–08	2006–07	2007–08	2006–07	2007–08
All States	√	√	√	√	√	-	-	-
Alaska	n/a	n/a	-	-	-	-	-	-
Arizona	-	√	-	√	-	√	-	√
Arkansas	√	-	√	-	√	√	√	√
Delaware	-	-	-	-	-	-	-	√
Florida	√	√	√	√	√	√	√	-
Iowa	√	√	√	√	√	-	√	-
North Carolina	√	-	√	-	√	-	√	-
Ohio	n/a	√	n/a	√	n/a	√	n/a	√
Tennessee	√	√	√	√	√	√	√	√

Exhibit reads: For all states in the 2007–08 school year, pilot growth models reduced the AYP gap between schools identified for improvement or under corrective action and schools not so identified; reduced the AYP gap between high-poverty and low-poverty schools; did not reduce the AYP gap between high-minority and low-minority schools; and did not reduce the AYP gap between urban and suburban schools. Note: "√" indicates the AYP gap was reduced, while "-" indicates that the AYP gap was not reduced.

Source: U.S. Department of Education, ED*Facts* and the Delaware state department of education.

Looking first at schools identified for improvement under *ESEA* for 2007–08, the percentage increase in such schools making AYP due to the growth models was greater than among schools not identified for improvement in Arizona, Iowa, and Ohio. The GMPP favored schools not identified for improvement in Arkansas and Delaware but did not favor either type of school in Florida and Tennessee. Overall, schools identified for improvement benefitted from the pilot program relative to schools not identified in both 2006–07 and 2007–08, though the effect differed between years for Arizona, Arkansas, and North Carolina.

Turning to school demographic characteristics, the effect of growth models on AYP was greater among schools enrolling higher proportions of students from poverty-level households in both 2006–07 and 2007–08. The pilot program was particularly effective in reducing the gap between the percent of low- and high-poverty schools making AYP in Florida, Iowa, Ohio, and Tennessee. Once again the effect differed between years for Arizona, Arkansas, and North Carolina. The growth component also was more likely to identify schools as making AYP among those with higher concentrations of minority students compared to schools with lower concentrations in the majority of GMPP states, though the effect differed by school year across all pilot states and for Iowa and North Carolina. The growth component did not benefit urban relative to suburban schools in either 2006–07 or 2007–08 but did close the gap in the majority of states in both years, though the effect differed across years for the majority of pilot states.

IV. Characteristics of Growth Models Affecting AYP Outcomes

The results presented in Chapters II and III show several large differences among the pilot states in the proportions of schools making AYP by growth as well as by status and safe-harbor. These differences are likely to arise from a number of sources, including features of the states' assessment systems, methods of determining AYP under status and safe-harbor, and characteristics of the growth models themselves. This chapter focuses on technical features of the growth models and their effect on AYP outcomes across the nine pilot states.

The school-level AYP determinations analyzed in Chapters II and III are aggregates of AYP determinations for subgroups, and AYP determinations for subgroups are aggregates of status and growth determinations for individual students. The growth model results for individual students were not collected for the ED*Facts* repository. However, the GMPP states were required to compile growth model results for all students in the participating grade levels. Specifically, each of the pilot states included in this Final Report provided student-level data with variables indicating (1) whether the student scored at the proficient level or higher on the state's reading and mathematics tests, and (2) whether the student was on-track to attain (or maintain) proficiency in the two subjects according to the growth model. In addition, some states were also able to provide scale score results for each student.

The analyses presented in this chapter draw on these student-level data to address several questions related to how and whether AYP determinations would be affected if the data collected by the GMPP states were used in various different ways. Six specific questions are addressed:

1. How do the nine state models compare in terms of the number of students they classify as on-track to proficiency?
2. How would the number of schools identified as making AYP by growth change in the nine pilot states if the growth model results were applied before status and safe-harbor criteria were used to assess AYP?
3. How do the three main types of growth models (transition matrix, trajectory, and projection) compare in terms of the number of students they identify as on-track and the number of schools that meet their AMO by growth criteria?
4. How do the main types of growth models compare in terms of their predictive accuracy of on-track students actually attaining proficiency?
5. How would AYP results change if the growth standards were not tied to students attaining proficiency standards?
6. To what extent did the pilot states have test scores for two or more years for all students, and did the non-matched students differ from their matched counterparts?

Comparison of Student On-Track and Proficiency Results

As discussed in Chapter I with respect to schools, it is theoretically possible for a student to have different proficiency determinations for growth and status. These possibilities are illustrated in

Exhibit 28, which outlines the four possible patterns. A student can be proficient under status and on-track to proficiency per the growth model (cell "A"), on-track to proficiency under the growth model but not proficient (cell "B"), proficient under status but not on-track to maintain proficiency (cell "C"), or neither proficient under status nor on-track to proficiency (cell "D"). The primary goal of the GMPP growth models is to identify students in cell B and to allow the schools to count those students the same as proficient students for accountability purposes. Although growth models allow for the identification of students in cell C as well, these students did not contribute negatively to AYP decisions for any school under the GMPP.

Exhibit 28
Conceptual Map of How Growth Model On-Track to Proficiency Designations Compare With Status Model Proficiency Designations for Students

Student On-Track to Proficiency Under the Growth-Only Model	Student Proficiency Under the Status Model	
	Proficient	Not Proficient
On-Track to Reach or Maintain Proficiency	A	B
Not On-Track to Reach or Maintain Proficiency	C	D

The percentages of students in each cell are shown in Exhibit 29. The percentages in cell A (proficient or higher and on-track to maintain or exceed proficiency) range from lows of 44 percent in Florida, 48 percent in North Carolina, and 49 percent in Arkansas to about two-thirds in Alaska, Delaware, Iowa, and Ohio and as high as 89 percent in Tennessee. The percentages of students classified as proficient but not on-track to maintain that level (cell C) divide into five states with 1 percent or fewer and four states with 7–9 percent. The groups largely coincide with the types of growth models (described in Chapter I) implemented: the states using transition matrix (Delaware and Iowa) and projection models (Ohio and Tennessee) have 0 percent in cell C while all of the states using trajectory models except Alaska have non-zero percentages. Alaska and the two transition matrix states opted to categorically define all proficient students as also on-track to proficiency. The projection model states did not impose such a rule but, as will be developed in the section of this chapter comparing generic versions of the models, their methods of determining whether students were on-track have a predictable tendency to classify proficient students as on-track. This tendency was amplified in Ohio by its procedure of adjusting upward the projected scores by two standard error units.

The main expectation is that the growth model will identify some students in cell B (not proficient but on-track) and thus increase the percentage of students that contribute to positive AYP determinations (the sum of cells A and B). Comparing the percentages of students in cell B, all states except Ohio had less than 10 percent of their students in this category. The variations among states in the cell B percentages are likely to reflect a combination of factors, including the effectiveness of the states' systems in moving below-proficient students to the on-track level as well as differences in the pool of eligible (non-proficient) students and the assessment instruments and relative difficulty of meeting the proficiency standards.

The cell B results can be standardized to some extent by viewing them in relation to the total pool of non-proficient students (the sum of cells B and D). Dividing the cell B percentage by the sum of cells B and D gives the rate at which non-proficient students are classified as on-track. This rate was by far the highest in Ohio where 75 percent of the non-proficient students were classified as on-track. The next highest rate was 45 percent in Tennessee, the other state using a projection model. Beyond those two states, Arizona had the next highest rate at 25 percent followed by Delaware and North Carolina at 18 percent, Florida at 13 percent, Arkansas and Iowa at 9 percent, and Alaska at 6 percent.

While Ohio had both the highest percentage in cell B and the highest rate of schools making AYP by growth documented in Chapter II, the relative sizes of the cell B percentages in the other states do not exactly correspond to their percentages of schools that made AYP by growth. Arizona and North Carolina, for example, had less than 1 percent of their schools classified as making AYP by growth despite having relatively high rates of non-proficient students identified as on-track to proficiency. The lack of correlation between the student and school results generally means the states differed in how the cell B students figured into the AYP determinations. For example, the pilot states differed in how they used confidence intervals and safe-harbor provisions, such that the impact of adding on-track students on AYP outcomes varied from state to state.

Exhibit 29
Distribution of Students According to How Their Proficiency and On-Track to Proficiency Classifications Compare, 2007–08

	Proficient and On-Track (Cell A in Exhibit 25)	Not Proficient but On-Track (Cell B in Exhibit 25)	Proficient but Not On-Track (Cell C in Exhibit 25)	Neither Proficient nor On-Track (Cell D in Exhibit 25)	Number of Students
Alaska	68%	2%	0%*	31%	68,787
Arizona	56%	9%	9%	27%	288,380
Arkansas	49%	4%	8%	39%	161,962
Delaware	67%	6%	0%*	27%	71,641
Florida	44%	6%	8%	42%	1,446,886
Iowa	68%	3%	0%*	29%	146,983
North Carolina	48%	8%	7%	37%	412,072
Ohio	68%	24%	0%	8%	584,988
Tennessee	89%	5%	0%	6%	401,659

Exhibit reads: Of the 68,787 students in the state of Alaska, 68 percent were proficient and on track to remain at or above the proficiency cut point, 2 percent were not proficient but were on-track to reach proficiency, 0 percent were proficient but not on-track to remain at or above the proficiency cut point, and 31 percent were neither proficient nor on track to proficiency.

* In the models employed by Alaska, Delaware, and Iowa, all proficient students are defined as being on-track.

Source: U.S. Department of Education, ED*Facts* and the Alaska, Arizona, Arkansas, Delaware, Florida, Iowa, North Carolina, Ohio, and Tennessee state departments of education.

Effects of Order of Application of Growth, Status, and Safe-Harbor

As discussed in Chapter I, the pilot states had the option to apply growth criteria to school AYP determinations in a number of ways: after status and safe-harbor, before status and safe-harbor, without status or safe-harbor, etc. Of the nine pilot states in 2007–08, all but Delaware elected to apply growth criteria after status and safe-harbor. Delaware, in contrast, applied growth criteria first, followed by status and then safe-harbor. If all three criteria (status, safe-harbor, and growth) are applied, the total number of schools making AYP is not affected by changes in the order in which the criteria are applied. Nonetheless, the quality of the information about the schools provided by the AYP data is arguably affected by the order of application, depending on the manner in which the AYP data are reported. This analysis assesses the extent to which changes in order of application would lead to states having higher rates of schools classified as making AYP by their growth models.

Theoretically, schools that would make AYP by growth-only would be more likely to attain or maintain AYP in the future than schools that make AYP by status but not growth, and also more likely to attain or maintain AYP in the future than schools that make AYP by safe-harbor but not by status or growth. If the AYP results are reported in a way that shows the criteria under which a school first achieves the AMO benchmark, applying growth criteria before status and safe-harbor could thus provide policymakers and the public with better information about school performance even though the overall numbers of schools making AYP would be the same. This is significant in light of policymakers' interests in possibly making greater use of student growth data, for it will identify schools that would have made AYP if growth were used not just before but also instead of status or safe-harbor.

The results in Chapter II indicated that, in most of the pilot states, relatively few additional schools made AYP by growth after status and safe-harbor criteria were applied. The logical possibilities of how hypothetical AMO results based solely on the on-track to proficiency data can intersect with actual AYP designations are shown in Exhibit 30. Referring to the cells in this table, the analysis in Chapter II focused on the number of schools that made AYP by growth (cells C plus G). However, those results do not provide a comprehensive account of the growth model data collected through the GMPP. First, it was possible for the schools that made AYP by status or by safe-harbor to also have had sufficient numbers of students classified by the growth model as on-track to proficiency to also meet their AMOs strictly under the growth criteria. Those schools would be located in cells A and B in Exhibit 30. Second, it was also possible for the schools that were classified as having made AYP by status-plus-growth to either meet (cell C) or not meet (cell G) the AMO if the students' on-track to proficiency data were used exclusively. Not meeting the AMO in this case would mean that a sufficient number of students testing as proficient (or above) are classified as non-proficient because they are not on-track to maintain proficiency.

Exhibit 30
How the Growth Model On-Track-to-Proficiency Designations Can Compare With AYP Designations for Schools

School AMO Designations Under a Hypothetical Growth-Only Model	School AYP Designations in *ESEA* Reporting, Based on Status-Plus-Growth Determinations			
	Made AYP by Status	**Made AYP by Safe-Harbor**	**Made AYP by Growth**	**Did not Make AYP**
Met AMO with growth-only	A	B	C	D
Did not meet AMO with growth-only	E	F	G	H

As noted in Chapter I, the ED*Facts* data classified schools in terms of AYP (and reporting groups in terms of AMO) with sets of mutually exclusive categories that do not allow one to address hypothetical questions of what would happen if the student growth model results were used differently. In order to assess the possible magnitude of the cells in Exhibit 27, the best resource is the status and growth proficiency data provided in the student files compiled by each of the pilot states. These data do not provide sufficient information to implement the states' methodologies for determining whether schools made AYP, but they do contain information on the key dimensions of reading and mathematics proficiency and thus allow assessments of whether reporting groups and schools as wholes met their AMOs.

In Alaska, Delaware, and Iowa, all students who scored at proficiency or higher on the status criteria were automatically also classified as "on-track" to proficiency per their growth models. The other six states, in contrast, applied their growth models to all students and identified at least some proficient students who were not on-track to continue to score at or above the proficiency levels in later grades.[35] It is important to note that in Alaska, Delaware, and Iowa it is possible for schools that made AYP by status to not meet the AMO using the growth model on-track indicator, even though all students who are proficient or higher were automatically counted as on-track as well. This is because these states identified a number of schools as making AYP by status that did not meet their AMO but that were close enough to count as making AYP because of the use of confidence intervals and multiyear averaging.

Status AYP and Growth-only Results. The results shown in the first column of Exhibit 31 indicate that for all nine states combined, 62 percent of the schools that made AYP by status also met their reading and mathematics AMOs using just the growth criteria. The states varied considerably, ranging from only 46 percent in Arizona and 47 percent in Arkansas and North Carolina to 75 percent or more in Ohio, Delaware, and Tennessee.

[35] There were a few exceptions to this rule in that some states did not actually calculate on-track indicators for students in certain grades. In those grades, the analyses reported in Exhibits 31 and 32 used the proficiency indicators for students in lieu of the on-track data in the calculation of the "growth-only" rates. This was done (instead of excluding those students) in order to improve comparability with the ED*Facts* data used in these analyses. Including versus excluding was particularly consequential in Ohio, which did not apply the growth model to third-graders and which had over half of the schools that failed to make AYP per ED*Facts* do so because third graders did not meet their AMOs in reading or mathematics.

Exhibit 31
Percentage of Schools Meeting AMO Using Only the Growth Model On-Track Indicator, by Standard ED*Facts* AYP Classification and State, 2007–08

Pilot States	Making AYP by Status	Making AYP by Safe-Harbor	Making AYP by Growth	All Schools Making AYP	Not Making AYP
All Nine States	62%	28%	31%	51%	3%
Alaska	54%	4%	ND*	42%	7%
Arizona	46%	17%	13%	43%	5%
Arkansas	47%	64%	23%	56%	3%
Delaware	79%	ND*	20%	77%	24%
Florida	58%	16%	20%	45%	2%
Iowa	57%	30%	35%	54%	15%
North Carolina	47%	3%	ND*	23%	<1%
Ohio	75%	45%	34%	52%	<1%
Tennessee	81%	5%	32%	69%	7%

Exhibit reads: Across all nine states, 62 percent of the schools that were classified in ED*Facts* as making AYP by status also met their AMOs when only the growth model on-track to proficiency indicator was used. Twenty-eight percent of the schools that were classified in ED*Facts* as making AYP by safe-harbor also met their AMOs when only the growth model on-track-to-proficiency indicator was used, and 31 percent of the schools listed as making AYP by growth met their AMOs using strictly the growth indicator.

* "ND" stands for "not defined;" there were no schools in these cells.

Source: U.S. Department of Education, ED*Facts* and the Alaska, Arizona, Arkansas, Delaware, Florida, Iowa, North Carolina, Ohio, and Tennessee state departments of education.

Taken at face value, these results indicate that the great majority of schools that made AYP by status in Ohio, Delaware, and Tennessee would have also met their AMO using only the on-track to proficiency data, but only 46 to 58 percent of the schools in the other seven pilot states that made AYP by status would have met their AMO using growth-only.

These findings provide some evidence that despite the low rates of making AYP by growth found in most states and documented in Chapter II, many schools had sufficiently high rates of students being on-track to proficiency to meet their AMOs using growth only. Under the normal practice of applying status before growth, the large numbers of schools that could have met their AMOs by growth-only were obscured.

Safe-harbor AYP and Growth-only Results. The second column of numbers in Exhibit 31 shows the percentages of schools that, according to ED*Facts*, made AYP by safe-harbor which also would meet the AMO for reading and mathematics using only the student on-track to proficiency indicator. The safe-harbor provision of *ESEA* was designed to identify schools making progress toward meeting the AMO even though their reading or mathematics proficiency levels, or both, are still below the AMO. Specifically, safe-harbor recognizes schools and subgroups that have decreased the percentage of students scoring below proficiency thresholds by 10 percent or more from one year to the next (e.g., decreasing the percentage of non-proficient students from 30 percent to 27 percent) and also make progress on the "other academic indicator," typically the average daily attendance rate for elementary and middle schools and the graduation rate for high schools. Identifying progress among low-performing schools is, of course, also the intention of the growth models, but the growth models are distinguished by (1) basing progress estimates on longitudinal student-level data, and (2) evaluating progress in terms of whether it is sufficient to reach the AMO within a specific time frame (e.g., by the time the student completes grade 8). By virtue of both these distinguishing features, schools meeting their AMOs by growth should be performing better than schools meeting their AMOs by safe-harbor but not by growth. Nonetheless, all of the pilot states except Delaware applied safe-harbor provisions before growth criteria in their AYP determinations.

The results in Exhibit 31 indicate that only 28 percent of the safe-harbor schools in the eight states that had any safe-harbor schools also met their AMO using the growth criteria. The percentages did not exceed 30 percent in any states except Arkansas (64 percent) and Ohio (45 percent).

The relatively low level of overlap of safe-harbor and growth-only outcomes in most states suggests that safe-harbor often identified schools as improving when the growth model indicated otherwise. The growth-only percent of students on-track could be a better gauge of actual test score improvement than safe-harbor because the growth-only measure relies on the longitudinal student records rather than the percentages proficient in each year. This is not to say that the safe-harbor schools that would not have met AMOs using growth-only were not making improvements but only to indicate instead that they were not making as much improvement as the growth-only measure requires. Insofar as the growth criteria are more informative about school improvement, they could be usefully applied to AYP determinations before safe-harbor is applied.

The third column in Exhibit 31 shows the percentages of schools classified as making AYP by growth in ED*Facts* that also met their reading and mathematics AMOs using only the student on-track to proficiency indicator. It is possible for these "made by growth" schools to not reach their AMOs using only growth criteria because the ED*Facts* classifications were based on status-plus-growth criteria instead of growth-only criteria. More specifically, the growth-only criteria classify some students who scored at or above the proficiency cut score as not on-track to proficiency and this makes the percentage of on-track lower than the percentage of proficient-plus-on-track. The results indicate that only 31 percent of the "made by growth" schools would have met their AMOs if only the on-track to proficiency indicator were used.

The fifth column in Exhibit 31 illustrates the lack of agreement between school-level results reported in ED*Facts* and those obtained from aggregating the student data within this evaluation project. Theoretically, the only schools that should show up in this column would be ones that met their AMOs for all subgroups but who failed to make AYP because of other criteria including participation rates and "other academic indicator" (average daily attendance for elementary and middle schools) results. This expectation appears to be born out in the overall rate of just 3 percent and the low rates in most states. However, the high rates for Delaware (24 percent) and Iowa (15 percent) suggest that other factors are complicating matters for those states, either on the side of how they made school AYP determinations for ED*Facts* or on the side of how they calculated the student on-track indicators provided to the evaluation team, or both.

Exhibit 32 shows the results of applying the growth-only criterion before status and safe-harbor within a framework of mutually exclusive categories like the ones currently used in ED*Facts*. For these calculations, the schools identified in ED*Facts* as not making AYP under any method were automatically kept as "not making AYP" in Exhibit 32 and the growth-only results applied to the schools classified in ED*Facts* as making AYP by status, safe-harbor, or status-plus-growth. These figures show that 27 percent of all schools in the pilot states would meet AMOs using growth-only, while the percent making it strictly by status would fall to 13 percent. While the overall rate of schools not making AYP (46 percent) would be about the same as under the status-plus-growth framework, the use of growth criteria before the others gives a clearer picture of the extent to which schools are attaining growth-to-proficiency goals than the status-plus-growth method.

Exhibit 32
Percentage of Schools Meeting the AMO Using Growth-Only and Making AYP by Other Means After Growth-Only Is Applied, by State, 2007–08

Pilot States	Meeting AMO by Growth-Only	Making AYP by Status	Making AYP by Safe-Harbor	Making AYP by Status plus Growth	Not Making AYP	Total	Number of Schools
All Nine States	27%	13%	7%	6%	46%	100%	13,401
Alaska	25%	21%	14%	0%	41%	100%	490
Arizona	32%	37%	5%	<1%	25%	100%	1,406
Arkansas	35%	8%	15%	4%	38%	100%	890
Delaware	54%	14%	0%	2%	30%	100%	182
Florida	11%	7%	3%	4%	76%	100%	3,280
Iowa	36%	26%	3%	1%	33%	100%	1,059
North Carolina	7%	7%	17%	0%	69%	100%	1,842
Ohio	34%	7%	3%	22%	33%	100%	2,896
Tennessee	58%	14%	12%	1%	15%	100%	1,356

Exhibit reads: In all nine pilot states, 27 percent of the schools would have met their AMO using only the growth models' on-track to proficiency results. Thirteen percent of the schools would have made AYP by status after applying the growth-only determination, 7 percent would have made AYP only by safe-harbor, 6 percent would have made AYP by status-plus-growth, and 46 percent would not have made AYP.

Source: U.S. Department of Education, ED*Facts* and the Alaska, Arizona, Arkansas, Delaware, Florida, Iowa, North Carolina, Ohio, and Tennessee state departments of education.

Effects of Types of Models

As described in Chapter I, the GMPP states used a variety of growth models to determine whether students were on-track to maintain or attain proficiency. These models were classified into three general types: transition matrix models, trajectory models, and projection models. By requirement of the pilot project, these models had to conform to seven core principles, including universal proficiency by 2014 (see Exhibit 3). This section contrasts these three types of growth models by using simplified model formulations and applying them to a common data source. This approach controls many of the nongrowth factors that can confound cross-state differences and allows generalizable conclusions about growth model features. No judgment is made about which models are preferable, and, certainly, the universe of possible models is larger than these three types. Instead, this section describes contrasting model features as they may support the contrasting policy priorities of individual states.

One of the findings from the preceding chapters of this report is that a number of nongrowth-related policy factors can moderate the impact of growth models. These factors are numerous and include safe-harbor, confidence intervals, AMOs, and cut scores. Unconditional cross-state differences between growth model results may be as much a function of these factors as of the types of models. To isolate the effects of the models, this section describes "generic" versions of the transition matrix, trajectory, and projection models used in the pilot states and applies them to longitudinal, standardized data from a single state: North Carolina.

The section begins with a description of the dataset and the parameters of three generic models. These generic models were chosen to be both representative of the state models in practice yet simple enough to be generalizable. The section continues with a comparison of student-level classifications under these three generic models. This comparison is performed for both a status-plus-growth approach, in which only non-proficient students are growth-eligible, and a growth-only model, in which proficient and non-proficient students must both make adequate growth. The section provides explanations and interpretations of differences among the generic models in the numbers and types of students they identify as on-track. The final part of the analysis aggregates the student-level results from each of the generic growth models to contrast school AYP results under the three models.

The questions addressed in this section and brief summaries of the findings with respect to growth model types are outlined in Exhibit 33.

Exhibit 33
Summary of Research Questions About the Generic Types of Growth Models

Research Questions	Findings
Which model is most likely to identify non-proficient students as on-track?	The trajectory and transition matrix models are more likely than the projection model. The contrast increases with higher (more stringent) proficiency cut scores.
Which model is most likely to identify proficient students as on-track?	The projection model is more likely than the trajectory and transition matrix models.
Which model is most likely to reclassify a school from not making AYP under status to making AYP when growth results are added to status results?	The trajectory model is slightly more likely to reclassify schools from non-AYP to AYP under a status-plus-growth method. This contrast increases with higher (more stringent) proficiency cut scores.
Which model is most likely to classify a school as making AYP using a growth-only criterion?	The projection model is much more likely to classify schools as making AYP under a growth-only criterion. This contrast decreases with higher (more stringent) proficiency cut scores.
Under which model are students identified as on-track to proficiency most likely to eventually reach or exceed proficiency?	The projection model provides better predictions of who will eventually reach or exceed the proficiency cut score.
Under which model are interim growth expectations most clearly identified?	The trajectory and transition matrix models identify interim annual growth targets for each student. The projection model can be configured to specify targets for each student, but the process is less straightforward.

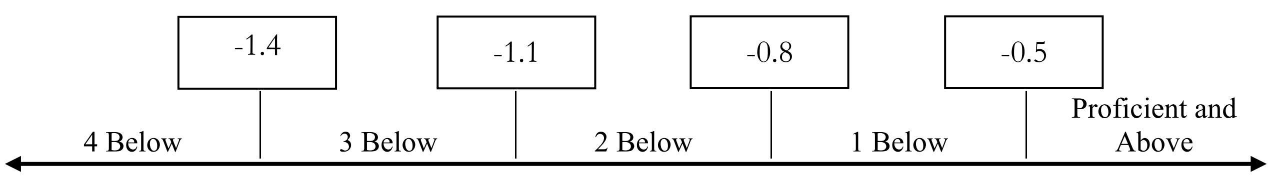
Generalizing From a Standardized North Carolina Dataset

The application of three types of growth models to a common dataset requires a dataset with particular properties. Three are listed here, each corresponding to one particular type of model. First, for projection models, at least five years of longitudinal data are required in order to estimate a projection equation with a time horizon of three years. The first two years may be used as predictors, and the fifth year is used as the outcome. In practice, more predictor years are desired, but two years of predictors are used here for cross-model comparability. The North Carolina database is the only readily available dataset among the pilot states with sufficient years to estimate regression coefficients.

Second, the trajectory model requires a vertical scale or some method of mapping cross-grade scores onto a common scale. For present purposes the scores are standardized to the so-called "z-scale," which has a mean of 0 and a standard deviation of 1 for each grade level. This is accomplished by taking each student's score, subtracting the state's average score in that student's grade, and dividing by the state's standard deviation of scores in that student's grade. An implicit assumption of this method of scaling is that students should roughly maintain their z-scores over time. Notably, this scaling approach allows for no substantive input about trends in student variation over time, no control over the expectation of whether high-scoring students should gain more than low-scoring students, and no input about whether student growth accelerates or decelerates over time. This atheoretical approach is thus inadvisable in practice but useful as a "generic" scaling approach.

Third, transition matrix models require multiple cut scores below proficiency and, for growth-only models, above proficiency. In practice, the average state proficiency rate is around 65 percent, and the median state proficiency rate is around 70 percent. To set representative cut scores on the z-scale, a proficiency cut score of -0.5 for each grade is used. Assuming scores that follow a standard normal distribution, the proficiency rate would be 69.14 percent for each grade. For the transition matrix model, which requires multiple cut scores below proficiency, equally spaced cut scores on the z-scale of -0.8, -1.1, and -1.4 are set as seen in Exhibit 34.

Exhibit 34
Illustration of Multiple Cut Scores for a Transition Matrix Model



In these four categories below proficiency, the percentages under normal distributions will be 9.7 percent, 7.6 percent, 5.5 percent, and 8.1 percent, for roughly equal proportions. These patterns are similar to those observed under transition matrix models in practice. Equally spaced cut scores are likely to be more generalizable than cut scores that generate equal proportions under normal assumptions, as there is no guarantee that normal assumptions will hold in practice or over time. Equally spaced cut scores are also easy to extend above proficiency. This extension will be necessary for a growth-only version of the transition matrix model described later.

To reiterate, this standardized North Carolina dataset bears little resemblance to the original North Carolina dataset. Its cut scores are not North Carolina's cut scores, and its trajectory model results will not follow the results reported in previous chapters. This was purposeful and results in a dataset with properties that generalize across states and support all three types of growth models. Certainly, growth model comparisons will be dependent on the scaling method and the cut scores in the data. When aggregating the student results to the school level for AYP purposes, systematic relationships between scales or cut scores and AYP results are also expected. Alternative cut scores than those shown in Exhibit 34 will be considered in a later section. More importantly, a later section introduces a general framework for visualizing the cut-score dependence of results.

Generic Growth Models

This section presents simplified, "generic" versions of growth models. These generic versions are designed to address two conflicting priorities: the comparability of results across models and the generalizability of results to actual state practices. To the extent possible, these models are designed so that differences can be interpreted as actual contrasts in model mechanics. To do this, cut scores, scaling decisions, and the available data are held constant across models while staying as true as possible to the operational models in growth model states.

This subsection presents status-plus-growth models, in which growth models have no effect on proficient students but classify some non-proficient students as "on-track." A student under a status-plus-growth model is either proficient, non-proficient and on-track, or non-proficient and not on-track. A later subsection presents growth-only models, in which the proficient cut score only functions as a target, and a student is characterized as "on-track" or "not on-track" regardless of current proficiency status.

The Generic Transition Matrix Model

The generic transition matrix model is similar to Delaware's model in that it has four categories below proficient. It is similar to Iowa's model in that any non-proficient student who has gained a category is considered on-track. Exhibit 35 displays the generic transition matrix model. Delaware's model can be represented similarly by inserting fractions in the above-diagonal, non-proficient cells of Exhibit 35, but the generic model is simpler and gives uniform credit to any gains (i.e., the cells with ones). Two years of longitudinal data are required to locate students in the table presented in Exhibit 35.

Exhibit 35
Student Weights Under the Generic Transition Matrix Model

		Year *t*+1				
		4 Below	**3 Below**	**2 Below**	**1 Below**	**Proficient**
Year *t*	4 Below	0	1	1	1	1
	3 Below	0	0	1	1	1
	2 Below	0	0	0	1	1
	1 Below	0	0	0	0	1
	Proficient	0	0	0	0	1

The Generic Trajectory Model

Trajectory models are used by Alaska, Arizona, Arkansas, Florida, and North Carolina and are described by Exhibit 9 in Chapter I. The generic trajectory model uses a linear trajectory from an examinee's first-year score to a proficient cut score four years in the future. A non-proficient examinee's second-year score must be at least one-quarter of the distance to the proficient cut score in order to be designated as "on track." Scores are standardized, and the proficient cut scores are set to be equal for all grades at -0.5. Thus, non-proficient students are on track if $X_2 - X_1 \geq (-.5 - X_1)/4$, where $X_2 - X_1$ is the gain over the first two years and $-.5 - X_1$ is the distance from the year 1 score to proficiency.

An equivalent expression of this model is that a student's gains, extended linearly for three years, must reach proficiency. This expression would be $X_2 + 3(X_2 - X_1) \geq -.5$, and, as expected, it is algebraically equivalent to the previous expression. Note that the first formulation establishes the initial year as the present-tense year with a four-year horizon, and the second formulation establishes the second, "current" year as the present-tense year with a three-year horizon. Because the latter formulation is more relevant to evaluating current year students, this model is described as an on-track-in-three model. Like the transition matrix model, this trajectory model requires only two years of data to evaluate students, as long as the cut score at the future time horizon is defined.

This generic trajectory model differs from the Arkansas model in that it is linear, but linearity is a simpler assumption and follows the other four trajectory model states. The generic model also differs from some models in that the time horizon from the initial year is four years. The effect of alternative time horizons is fairly straightforward to predict and is described in upcoming sections. Many state models have an "on-track in three years or by grade *x,* where *x* is determined by the state, whichever is sooner" policy. For the standardized North Carolina dataset with grades 3–8 and an assumed grade 8 graduation, this amounts to a policy where current eighth-graders must be proficient, current seventh-graders must be on-track in one year ($N = 1$), current sixth-graders must be on-track in two years ($N = 2$), and current fifth-graders and younger must be on-track in three years. The stricter time horizons for higher grades is a feature of the generic trajectory model.

The Generic Projection Model

The generic projection model is a simplified version of Ohio's and Tennessee's model. These simplifications achieve clear comparisons to the transition matrix and trajectory models. The simplified approach mimics the data use of transition matrix and trajectory models, which rely almost solely on the current year's data (grade g) and the previous year's data (grade g-1) to arrive at decisions. Similarly, for the simple projection model, students are treated as if they have only the most recent two years of reading and mathematics scores in order to make reading and mathematics on-track decisions, respectively, as seen in the following regression equations:

$$\widehat{R}_{g+N} = \overline{\mathrm{R}}_{g+N} + \hat{\beta}_{R1}\left(R_{g-1} - \overline{\mathrm{R}}_{g-1}\right) + \hat{\beta}_{R2}\left(R_g - \overline{\mathrm{R}}_g\right)$$

$$\widehat{M}_{g+N} = \overline{\mathrm{M}}_{g+N} + \hat{\beta}_{M1}\left(M_{g-1} - \overline{\mathrm{M}}_{g-1}\right) + \hat{\beta}_{M2}\left(M_g - \overline{\mathrm{M}}_g\right).$$

Here, $\widehat{R}_{g+N}$ and $\widehat{M}_{g+N}$ denote the predicted score for reading and mathematics in grade (g + N), respectively, N years from the current grade g. Similarly, R_i and M_i denote the reading and mathematics scores in grade i, and $\hat{\beta}_{jk}$ denotes the k^{th} regression coefficient for subject j. The regression coefficients cannot be estimated from the current cohort, as current students in grade g cannot know their scores in some future grade $(g + N)$. Thus, regression coefficients are estimated from a covariance matrix of test scores for a previous "reference" cohort that has completed the target grade level (g + N) of the state's growth model. The prediction is obtained by entering current student scores into a prediction equation that uses the previous cohort's regression coefficients. For these generic projection models, N is set at 3, projecting to a grade three years from the current grade, following the on-track-in-three trajectory model. Due to the use of a standardized, "generic" dataset, these equations simplify as follows:

$$\hat{R}_{g+N} = \hat{\beta}^*_{R1} R_{g-1} + \hat{\beta}^*_{R2} R_g$$

$$\widehat{M}_{g+N} = \hat{\beta}^*_{M1} M_{g-1} + \hat{\beta}^*_{M2} M_g$$

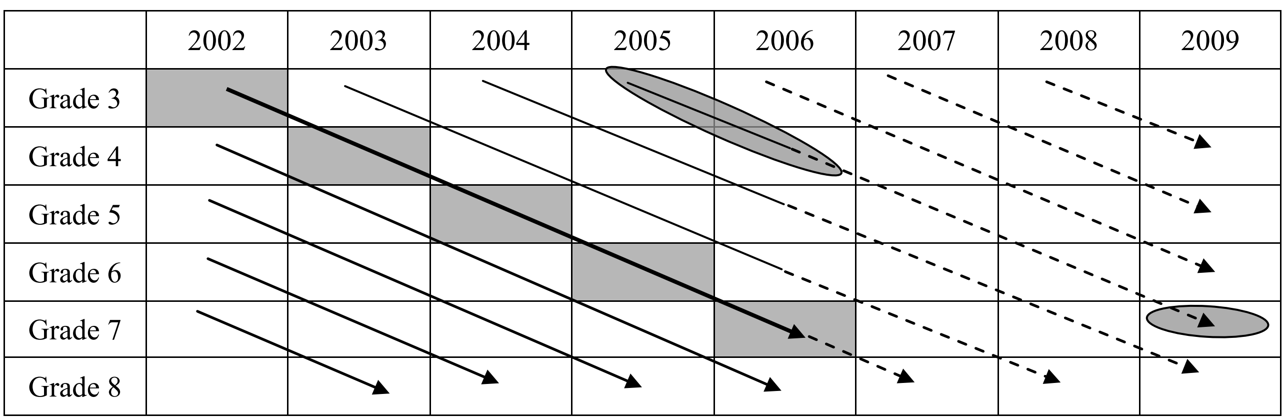
Here, the regression coefficients, $\hat{\beta}^*_{jk}$, are distinguished with a "$*$" to denote their estimation from a past reference cohort. In practice, projection models typically use more than two predictors (Ohio and Tennessee, for example, require a minimum of three prior scores for predictions[36]). However, additional years of data would confound comparisons to transition matrix and trajectory models, as classification discrepancies could be explained by these additional data. Additionally, as upcoming sections argue, the conclusions drawn from the simplified projection model generalize to models with greater numbers of predictors. Students who are missing scores in the most recent two years in a particular subject will not be eligible for the growth model in that subject. This follows the same missing data approach as transition matrix- and trajectory-model states.

[36] These details are discussed at http://www2.ed.gov/admins/lead/account/growthmodel/oh/ohadd2.doc (p. 8) for Ohio, and at http://www.quickanded.com/2009/03/tennessee-growth-models-response-from.html (first footnote) for Tennessee.

An important difference between the generic projection model developed here and the projection model used in Ohio is that no confidence interval is applied to the generic projected scores. As described in Chapter II, Ohio estimated a standard error for each student's projected score and added to the projected score a quantity equal to two standard error units; the augmented project score was then compared to the cut score for the target year in order to determine whether the student was on-track. The other pilot state using a projection model, Tennessee, did not adjust projected scores.

The available grades and years of the standardized North Carolina dataset are shown in Exhibit 36. The most recent year with full data availability was 2006; this is established as the "current year" for student growth classification. For a hypothetical fourth-grade cohort in 2006, transition matrix and trajectory models need only two years of test score data, as represented by the central dark gray oval, to establish whether students are on-track. In contrast, projection models require a five-year dataset to project from grade 3 and grade 4 data to proficiency in the target grade 7 (three years in the "future"), shown for the 2006 cohort as the gray oval in 2009. Because 2009 data are not available in the "current year" of 2006, a reference cohort is used to estimate coefficients as described previously. This cohort is indicated in the table with the bolded arrow in the gray shaded cells. Once the coefficients are estimated from the reference cohort, the data from the current cohort are entered into the equation, and predicted student scores in grade 7 are compared to the grade 7 proficiency cut score to determine on-track status. This procedure is a simplification of Tennessee and Ohio's approach that captures the fundamental regression-based features of the model while allowing contrasts to the other two model types.

Exhibit 36
Schematic Diagram of Data Required for Applying the Projection Model



Generic Model Extensions for Growth-Only Results

The generic trajectory model is extended easily to a "growth-only" approach by requiring that all students, whether proficient or not, be on track in three years or by a Grade 8 graduation. The same equations would apply to both proficient and non-proficient students. The projection model is likewise straightforward to extend: Both proficient and non-proficient students are entered into the prediction equation and must be predicted to be proficient in three years or by grade 8.

Transition matrix models require categories above the proficient cut score in order to extend inferences to proficient students who decline. A simple approach is to take the equally spaced cut scores below proficiency and extend them above proficiency. These cut scores follow, with the proficient cut score shaded:

-1.4	-1.1	-0.8	-0.5	-0.2	0.1	0.4	0.7	1.0	1.3

As before, a non-proficient student who does not change categories is not on-track. In contrast, the growth-only transition matrix model gives proficient students who do not change categories the benefit of the doubt and designates them as on-track.

Student-Level Results for Status-Plus-Growth Models

As shown in Exhibit 36 above, the 2006 cohort (academic year 2005–06) is the most recent representative year of data in the available North Carolina dataset. The analysis focused on all Grade 3–7 students in this year that also had data from the previous year. These "two-year match rates" range from 90 percent to 96 percent. The results will be referenced by the grade of the students in 2006, the year of focus. The Grade 5 results, for example, refer to the students in Grade 5 in 2006 for whom Grade 4 data in 2005 were available. Grade 8 students are not eligible for the growth model, and their results are not reviewed. The 2006 Grade 3 Math cohort was not included because of extremely low match rates to the "Grade 2.5" pretest in that particular year.

As described above, cut scores were set in a generic fashion by standardizing scores and selecting a proficiency cut score of -0.5 on the z-scale. (See the "Alternative Cut Scores" section below for discussion of alternative proficiency cut scores.) A z-score of -0.5 results in about 69 percent of the students classified as proficient with normally distributed data. For reference, Exhibit 37 shows that resulting hypothetical proficiency rates based on the -0.5 cut scores hover around 70 percent as expected, with variability arising from deviations from normality and the discrete nature of the test scores.

Exhibit 37
Overall Proficiency Rates for Reading and Mathematics for Each Grade in 2006

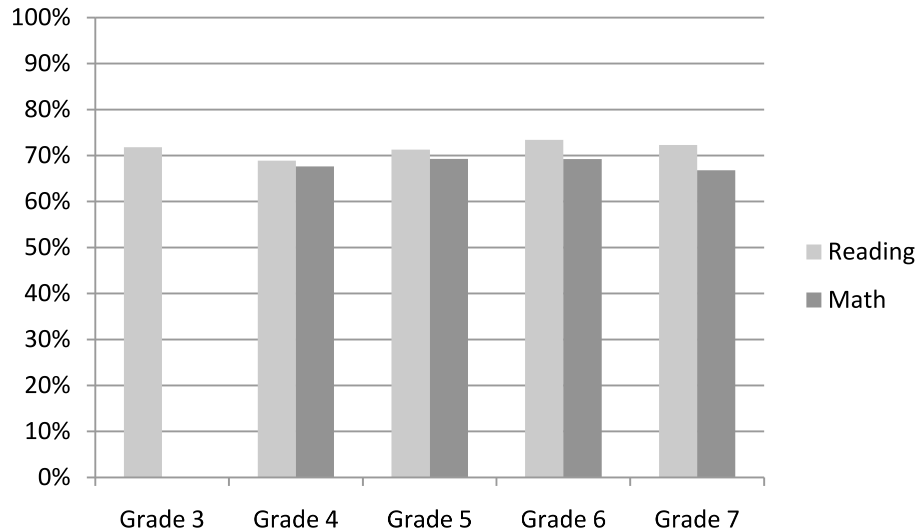
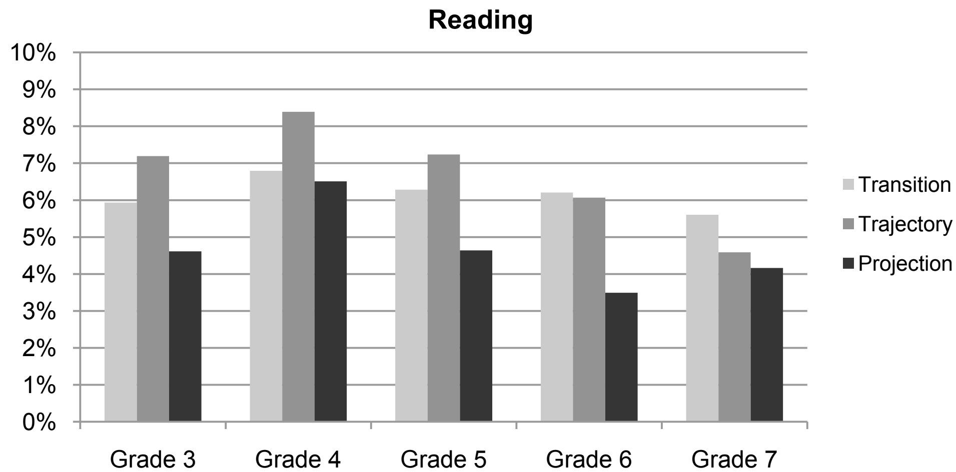


Exhibit reads: Of the students assessed in 2006, 71.8 percent of those in Grade 3 are classified as proficient with respect to reading. Mathematics was not assessed for Grade 3.

Exhibit 38 shows overall percentages of "on-track" students as classified by each model in each grade. The projection model always classifies fewer students as "on-track" and shows even greater relative stringency in math. For reading, the percentage of students classified as "on-track" ranges from about 3.5 percent to about 8.5 percent, and for math, from about 3 percent to 8 percent. The trajectory model becomes more stringent in Grade 6 and Grade 7, as is expected of a shortened time horizon to proficiency in Grade 8 (two years and one year respectively).

Exhibit 38
"On-Track" Classification Rates for All Students for Reading and Mathematics by Model and Grade in 2006



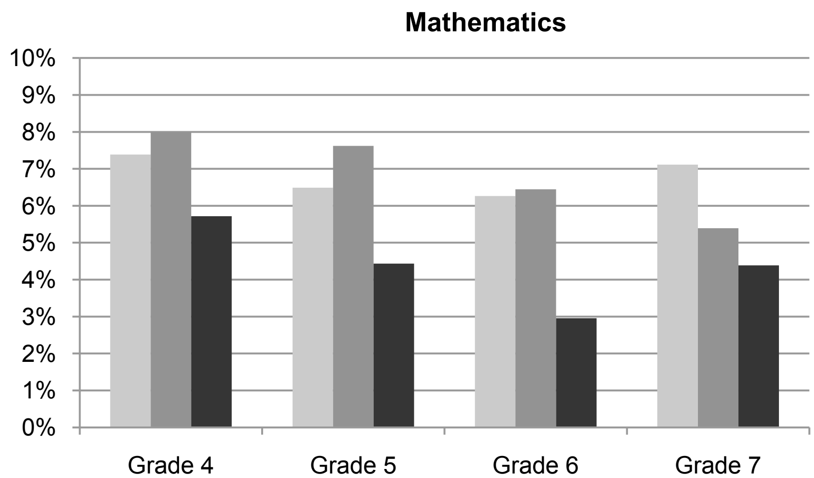
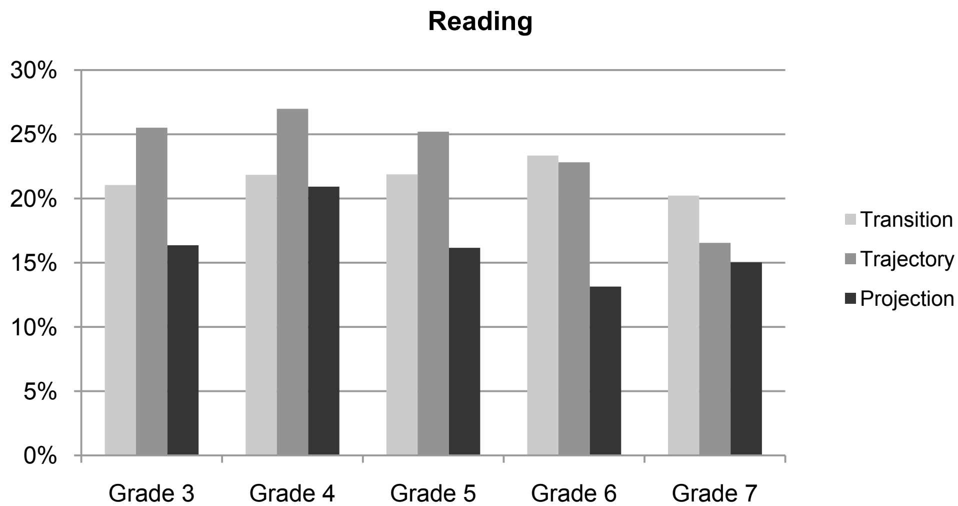


Exhibit reads: Of the third-grade students assessed in reading in 2006, the generic transition model classified 5.9 percent as on-track, while 7.2 percent and 4.6 percent were classified as on-track by the generic trajectory and projection models, respectively.

Exhibit 39 shows the same results but, because growth models can only make a difference for non-proficient students when the results are used in the typical status-plus-growth framework, the percentages are rescaled to represent the classification rates for eligible (non-proficient) students. Whereas Exhibit 38 shows the percentage of *all* students classified as on-track by the growth model, Exhibit 39 shows the percentage of *eligible* (i.e., non-proficient) students classified as on-track by the growth model. Between 10 percent and 27 percent of eligible students are classified as on-track by the three different growth models. The same pattern in relative magnitude among the three models as seen in Exhibit 38 is also apparent in Exhibit 39 for just the eligible students: The trajectory model tends to classify the most students as on-track, followed by the transition matrix model. Within each grade, the projection model classifies the fewest eligible students as on-track.

Exhibit 39
"On-Track" Classification Rates for Eligible (Non-Proficient) Students for Reading and Mathematics by Model and Grade in 2006



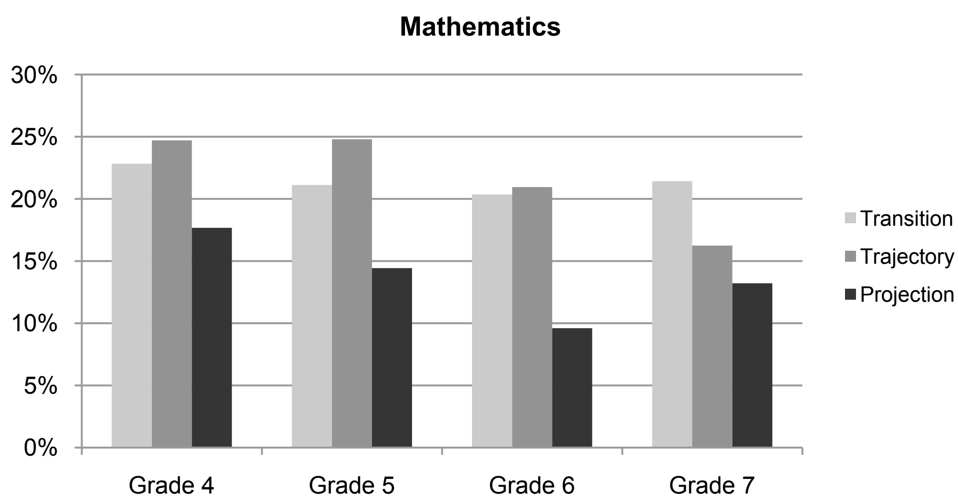


Exhibit reads: Of the eligible (non-proficient) students assessed in 2006, 21.0 percent of those in Grade 3 are classified students as "on-track" with respect to reading by the transition model.

It is noteworthy that the generic results are at odds with the empirical results shown in Exhibit 29, in which the two states using projection models had the highest rates of non-proficient students classified as on-track (75 percent in Ohio and 45 percent in Tennessee, compared with 25 percent or less in the other pilot states). As noted in Chapter II and in the section "Comparison of Student On-track and Proficiency Results" earlier in this chapter, much of the high on-track rate observed in Ohio is likely explained by its procedure of augmenting the projected scores by two standard error units. However, an explanation for the high rate for Tennessee is not readily available. Two possible factors are listed here and discussed more fully in the upcoming sections. First, the model interacts with the difficulty of cut scores, both generally and across grades; this is discussed more fully below in the section "Alternative Cut Scores for Status-Plus-Growth Model." As noted in that discussion, the projection model is particularly sensitive to where cut scores are located, with higher rates of non-proficient students identified as on-track when cut scores are lower. In 2007–08, Tennessee set relatively low cut scores and had one of the highest proficiency rates in the nation. Although these cut scores have changed, this may have contributed to the relatively high rate of on-track students observed in Exhibit 29. A second, less likely explanation is that the covariance structure of the Tennessee data differs dramatically from the covariance structure of the North Carolina data. However, cross-grade covariance relationships do not generally differ across states to an extent that explains the difference between Exhibits 29 and 39.

Exhibit 40 shows the pairwise percent agreement across models. Because all models classify proficient students as proficient, this is not counted as "agreement," so only non-proficient students make up the denominator of the percent agreement statistic. The findings suggest that the transition matrix and trajectory models have similar mechanisms and identify greater proportions of students as "on-track" than projection models. Transition matrix and trajectory models agree on over 90 percent of growth-eligible (non-proficient) students. Their agreement rates with projection models are closer to 60 percent. This agreement is wholly accounted for by common classification of not on-track students, with the exceptions of Grade 4 Reading and Grade 7 Math, which have a handful of students who are commonly classified as on-track. Agreement about on-track students is near zero. That is, if a transition matrix or trajectory model classifies a student as making adequate growth, the projection model almost always disagrees, and vice versa. The theoretical reasons for this are laid out in the section below, entitled "A Framework for Comparing the Models."

Exhibit 40
Pairwise Percent Agreement of Models for Eligible (Non-Proficient) Students, Reading and Mathematics

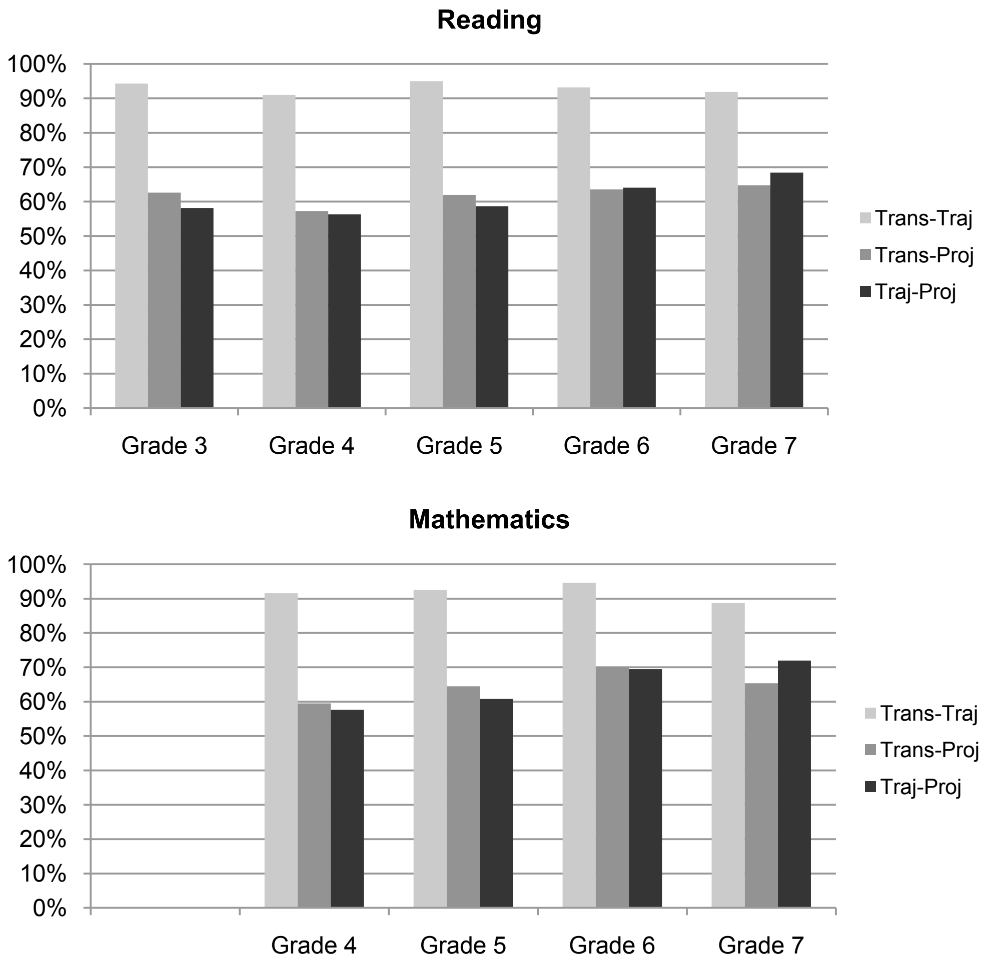


Exhibit reads: The transition matrix (Trans) and trajectory (Traj) models agreed on 94.3 percent of eligible (non-proficient) Grade 3 students assessed with respect to reading. The transition matrix and projection (Proj) models agreed on 62.6 percent, while the trajectory and projection models agreed on 58.1 percent in reading.

As noted previously, the generic projection model differs more from its real-life counterpart than the transition matrix and trajectory models in that this projection model uses only one subject and only two years of prediction. This was done to ensure that all growth models were using similar data. The projection model can be extended across subjects and grades, but this will not necessarily increase the agreement rates. The reasons for this are also laid out in the section "A Framework for Comparing the Models" below.

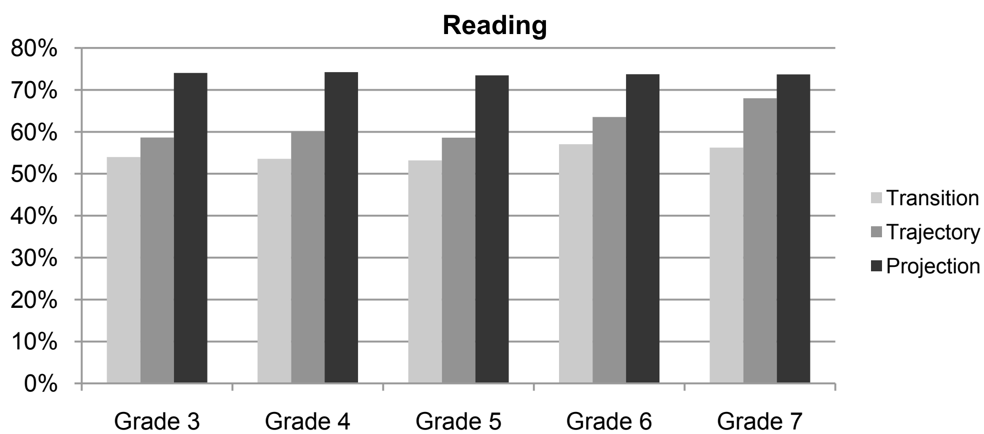
Student-Level Results for Growth-Only Models

This section overviews the results from a growth-only approach. This parallels the analyses at the beginning of this chapter in which growth-only criteria were applied to each GMPP state.

The results here also parallel the immediately prior status-plus-growth section, in which all models agreed about proficient students, leaving model contrasts centered on non-proficient students. In this section, students are either on-track or not on-track, and models can theoretically disagree about any student.

Exhibit 41, below, is the analog of Exhibit 39 from the status-plus-growth analysis. Exhibit 41 shows the "on-track" classification rates for all students by model and grade. Transition matrix models classify the fewest students as on-track: between 53 percent and 57 percent. Trajectory models classify more students as on-track: between 59 percent and 68 percent. The higher percentages are in the higher grades. The two-year horizon in Grade 6 and the one-year horizon in Grade 7 are less lenient for non-proficient students but more lenient for proficient students. Because there are more proficient students, this manifests as an increase in the overall on-track classification rate. Projection models have the highest classification rates: between 70 percent and 74 percent and fairly uniform across grades.

Exhibit 41
"On-Track" Classification Rates Using Growth-Only Results for All Students for Reading and Mathematics, by Model and Grade in 2006



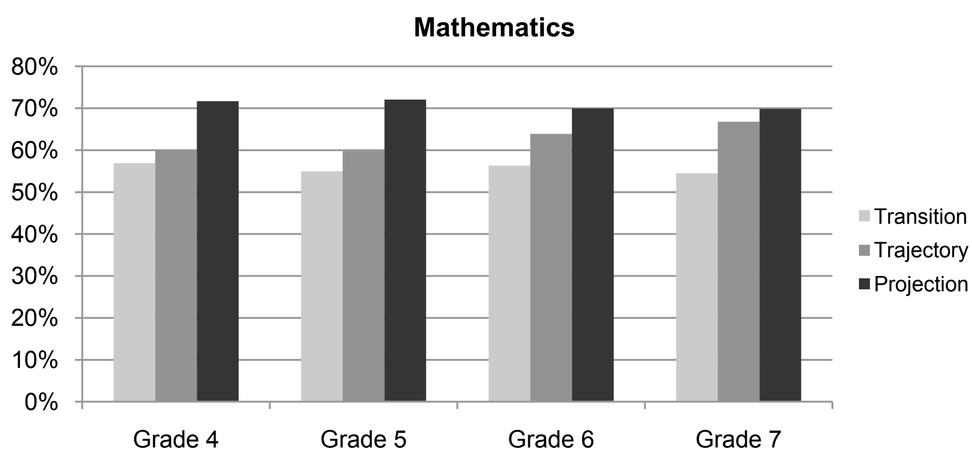
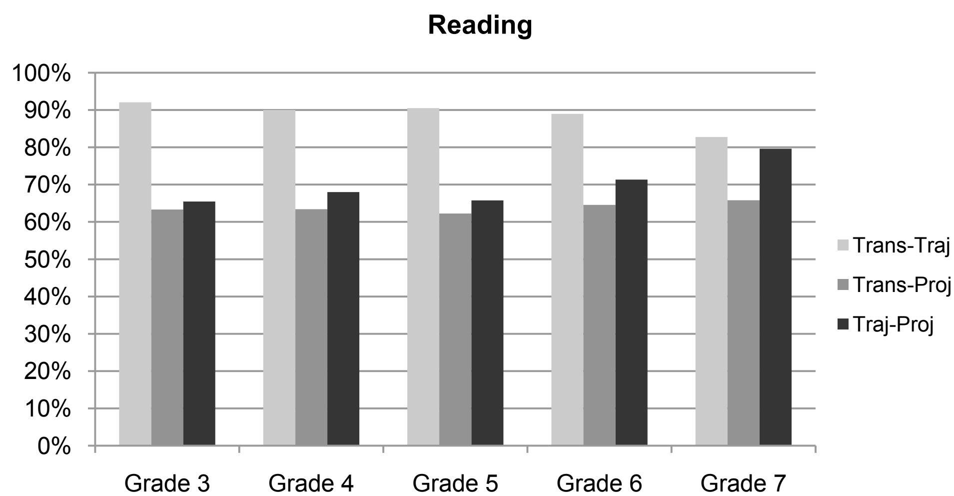


Exhibit reads: Of the students assessed in 2006, 54.0 percent of those in Grade 3 are classified as "on-track" with respect to reading by the transition model.

Exhibit 42 shows the pairwise percent agreement of the models, defined as the percentage of students for whom on-track classifications are agreed upon by a pair of models. This is the analog of Exhibit 40 from the status-plus-growth analysis. Transition matrix and trajectory models are again the most similar in overall function, hovering around 90 percent agreement for the lower grades and decreasing in the higher grades to near 82 percent. This decline in the higher grades is explained by the decreasing time horizon and can be visualized in the upcoming section "A Framework for Comparing the Models."

Exhibit 42
Pairwise Percent Agreement of Models Using Growth-Only Results for All Students in Reading and Mathematics in 2006



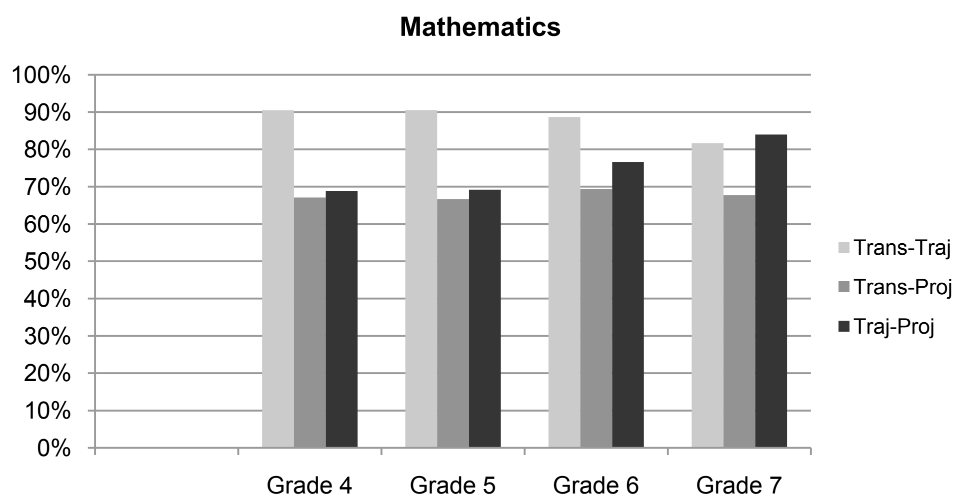


Exhibit reads: The transition matrix (Trans) and trajectory (Traj) models agreed on 92.1 percent of Grade 3 students assessed with respect to reading.

Exhibit 42 also shows that the lowest agreement in each grade is between the transition matrix and projection models: between 62 percent and 69 percent. Exhibit 42 shows similarly low agreement between the trajectory model and the projection model. The agreement rates are 65 percent, but they increase to 80 percent and 85 percent in Grade 7. This dramatic increase in agreement is largely explained by the increased leniency of the trajectory model on proficient students due to the time horizon. The explanatory framework developed below illustrates why this agreement between the trajectory and projection models increases with convergence toward the end of the growth model's grade span.

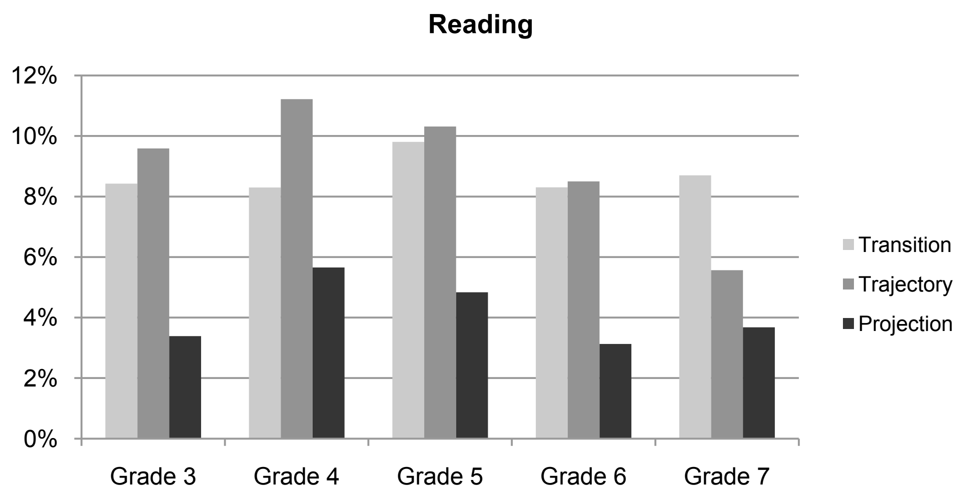
Alternative Cut Scores for Status-Plus-Growth Models

One important way in which states differ is in the setting of cut scores, with some states setting higher cut scores and typically posting lower proficiency rates while others set lower cut scores and classify more students as proficient. In this section, we briefly examine results of using the same data but raising the proficiency cut score from -0.50 to -0.25. For these data, this results in proficiency rates near 60 percent compared to the previous proficiency rates nearer to 69 percent. To adapt to this change, the transition matrix model cut scores below proficiency were also raised slightly and spaced at 0.35 intervals instead of 0.3.

Exhibit 43 shows that transition matrix and trajectory models classify larger proportions of students (compare to Exhibit 38), because there are more non-proficient students to classify. In contrast, the projection model classifies slightly fewer students. The raised cut scores leave more non-proficient students to classify, but the classification approach of projection models, which contrasts so strongly with the other two models, interacts differently with the bivariate distribution of scores. This can be visualized in the upcoming section, "A Framework for Comparing the Models."[37]

[37] The cut-score dependence of growth models has been described by A.D. Ho, D.M. Lewis, and J.L.M. Farris (2009) The dependence of growth-model results on proficiency cut scores. *Educational Measurement: Issues and Practice*, 28 (4), 15–26.

Exhibit 43
"On-Track" Classification Rates for All Students for Reading and Mathematics by Model and Grade in 2006: Higher Cut Scores, Proficiency Rates Near 60 Percent



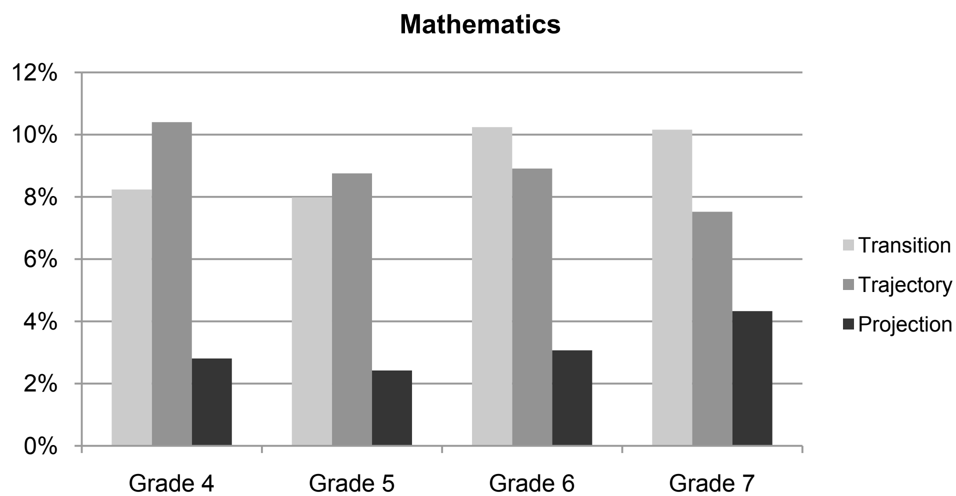
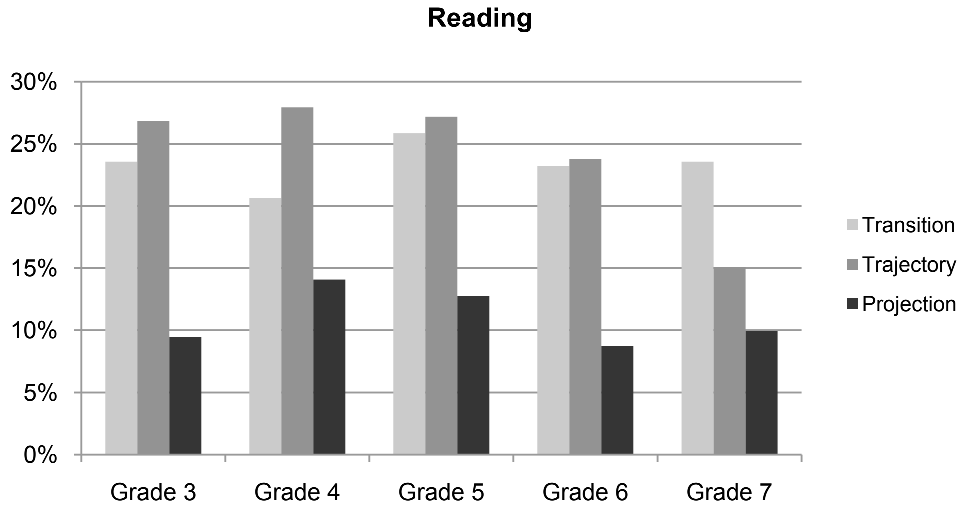


Exhibit reads: Of the students assessed in 2006, 8.4 percent of those in Grade 3 are classified as "on-track" with respect to reading by the transition model when higher cut scores are used.

Exhibit 44 shows the proportion of eligible students classified by the models. These values are generally similar to Exhibit 39 with the exception of the projection model identification rates, which have declined substantially. As cut scores increase and proficiency rates decrease, the relative impact of projection models will decline. This is also visible in the upcoming framework in the following section.

Exhibit 44
"On-Track" Classification Rates for Eligible (Non-Proficient) Students for Reading and Mathematics by Model and Grade in 2006: Higher Cut Scores, Proficiency Rates Near 60 Percent



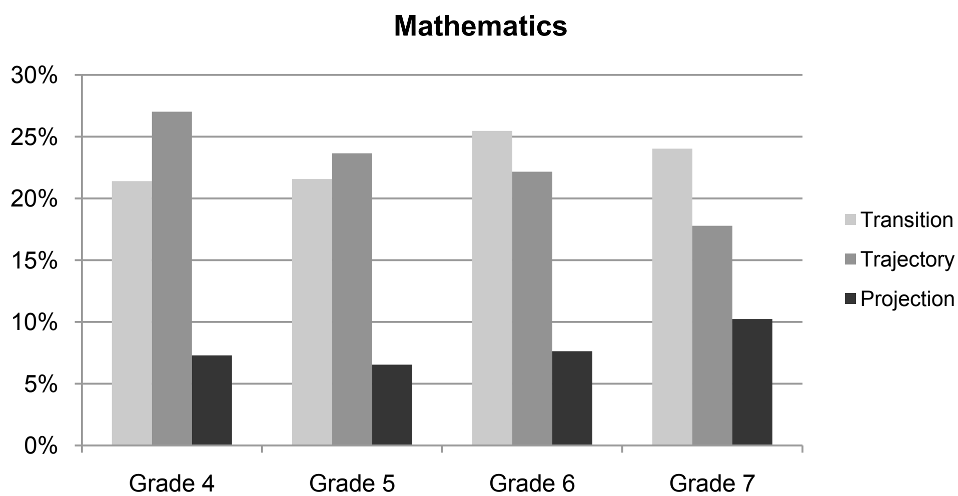


Exhibit reads: Of the eligible (non-proficient) students assessed in 2006, 23.6 percent of those in Grade 3 are classified as "on-track" with respect to reading by the transition model when higher cut scores are used.

A Framework for Comparing the Models

Differences and disagreements in growth model classification can be visualized with a bivariate framework (modified from Ho, Lewis, and Farris, 2009). The framework allows visualization of the kinds of students classified as on-track by each of the three models and explains the bar-plot exhibits from the preceding sections in a more coherent structure. The framework is somewhat complex, but it affords the following:

- Visualization of the transition matrix model as a categorical approximation of the trajectory model.
- Visualization of the projection model as a stark contrast to the trajectory and transition matrix models.
- Visualization of the projection model as a model that rewards high average achievers, whereas the trajectory and transition matrix models can be visualized as models that reward students making gains.
- Visualization of the projection model's relative preference for previously high-scoring decliners and both the transition matrix and trajectory models' relative preferences for low-scoring gainers, albeit those that might be less likely to actually reach proficiency.
- Visualization of the projection model's relatively low classification rates for status-plus-growth models but relatively high classification rates for growth-only models.
- Visualization of how the trajectory model's classification rate over the projection model will increase with increasing cut scores.
- Visualization of how shorter time horizons increase the overlap between trajectory and projection models.

Exhibit 45, on the following page, displays lines that overlay an imagined scatterplot of student scores, with a current year's scores plotted against a past year's scores. The scores are not shown to prevent a cluttering of the diagram, but they may be imagined as a cloud centered on the point (0,0) in the upper right portion of the plot and stretching as an ellipse from the lower left to the upper right due to the correlated nature of adjacent-grade scores. In this context, the y-axis represents the 2006 scores for a given grade cohort, say, Grade 5, and the x-axis represents the 2005 scores for the same cohort, in this case, in Grade 4. The scores are standardized within both years to a mean of 0 and a standard deviation of 1, so axis values can be interpreted on the z-scale. The framework intentionally focuses on lower-scoring students, and the figure is bounded by -2 and 0.5. Important features of the plot include the main diagonal, along which students have the same z-scores in both years. Students above this diagonal are increasing their scores (in standard deviation units from the mean) over time, and students below this diagonal are declining. The proficiency cut scores are set at z-scores of -0.5 in both 2005 and 2006, as described above. These are represented by a bold vertical and horizontal line respectively. Students above the horizontal "Proficient in 2006" line are proficient in the "current year" and are unaffected by status-plus-growth models.

The transition matrix model cut scores of -0.8, -1.1, and -1.4 are also shown.[38] Under the transition matrix model, non-proficient students who have gained a category are deemed "on-track." This can be visualized in Exhibit 45 by the shaded regions. All of the shaded regions lie above the main diagonal, as befits any increase in z-scores, and only non-proficient students are highlighted, as those above the -0.5 line are simply proficient. The staircase shape arises from

[38] The consequences of higher cut scores are generally predicted by the framework, which slides up the diagonal over the imagined bivariate scatterplot in Exhibit 45. The target areas illustrated by Exhibit 45 begin to incorporate more central and denser regions of the bivariate scatterplot.

the categorical nature of the transition matrix model, which does not distinguish score levels within any score category.

Exhibit 45
A Bivariate Framework Showing the Non-Proficient Students Deemed "On-Track" by a Transition Matrix Model

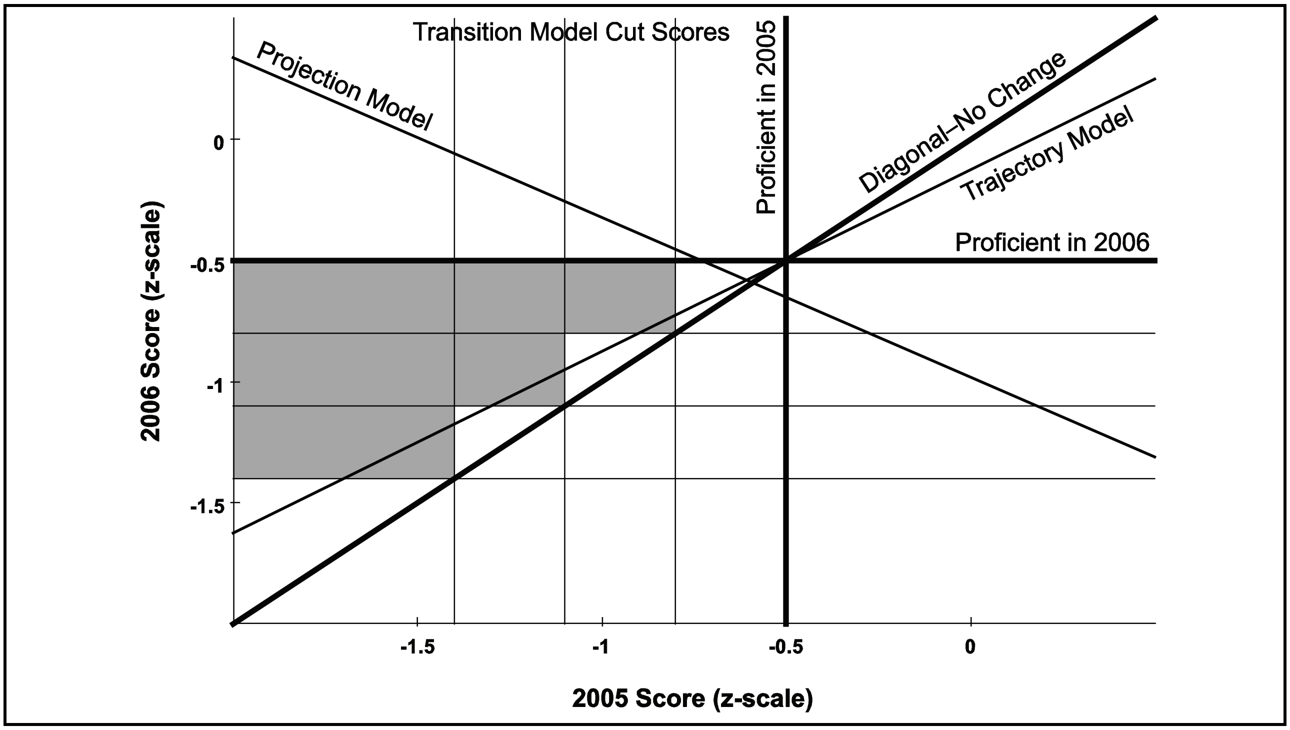


Exhibit reads: The shaded area corresponds to students who were not proficient in 2006 but whose growth from 2005 to 2006 classified them as on-track under the transition matrix model. Non-proficient students in 2006 classified as on-track by the trajectory model lie in the triangle (not shaded) above the Trajectory Model line and below the Proficient in 2006 line. Non-proficient students in 2006 classified as on-track by the projection model lie in the triangle (not shaded) above the Projection Model line and below the Proficient in 2006 line.

Exhibit 45 also includes lines for the other growth models. The trajectory model line arises from the trajectory model decision rule: If a student's score gains from 2005 (x) to 2006 (y) are continued for the next three years (or to eighth grade, whichever is sooner) and the student is then proficient or above, then that student is on-track. On this scatterplot, this amounts to the following equation, which is again algebraically equivalent to the generic trajectory model formulation presented in the previous section:

If $y < -0.5$ (non-proficient) AND $y + 3*(y-x) \geq -0.5$, then on-track.

These equations are both plotted as lines in Exhibit 45: the Proficient in 2006 line and the Trajectory Model line respectively. Students classified as "on-track" by the trajectory model will lie in the semi-infinite triangle (not shaded) above the Trajectory Model line and below the Proficient in 2006 line.

For later grades, the proficiency horizon may be less than three years, so the slope of the trajectory model line will decrease and pivot about the "proficiency origin" of (-0.5, -0.5). If all else is equal across grades, this will effectively decrease the proportion of students classified as "on-track" for Grade 6 students (who must be on-track in two years) and Grade 7 students (who must be on-track in one year). This decline can be seen in Exhibits 38 and 39.

Exhibit 46 shows areas of discrepancy between transition matrix and trajectory models. Students falling in the darker-shaded triangles will be classified as on-track by the trajectory model but not the transition matrix model. Students falling in the lighter-shaded triangles will be classified as on-track by the transition matrix model but not the trajectory model. These triangles are fairly small when all non-proficient students below the Proficient in 2006 line are considered. This accounts for the high agreement rates for the two models in Exhibit 40. Discrepant classifications are fairly balanced in this case.

Exhibit 46 uses Grade 5 data for an "on-track in three years" model. For Grades 6 and 7, in which the slope of the Trajectory Model line decreases, the discrepant classifications will be less balanced. Not only will the trajectory model begin to classify fewer students as on-track, but the discrepancies will begin to be more fully accounted for by the leniency of the transition matrix model. Exhibit 46 illustrates that the conceptual simplicity of the transition matrix models' discrete performance categories carries a cost of some misclassification of students compared with the trajectory model. However, the interpretability of the transition matrix model's student-level results may be more straightforward for teachers, students, and parents than score-based trajectory models despite the loss in classification accuracy.

Exhibit 46
Areas of Discrepancy Between Transition Matrix and Trajectory Models

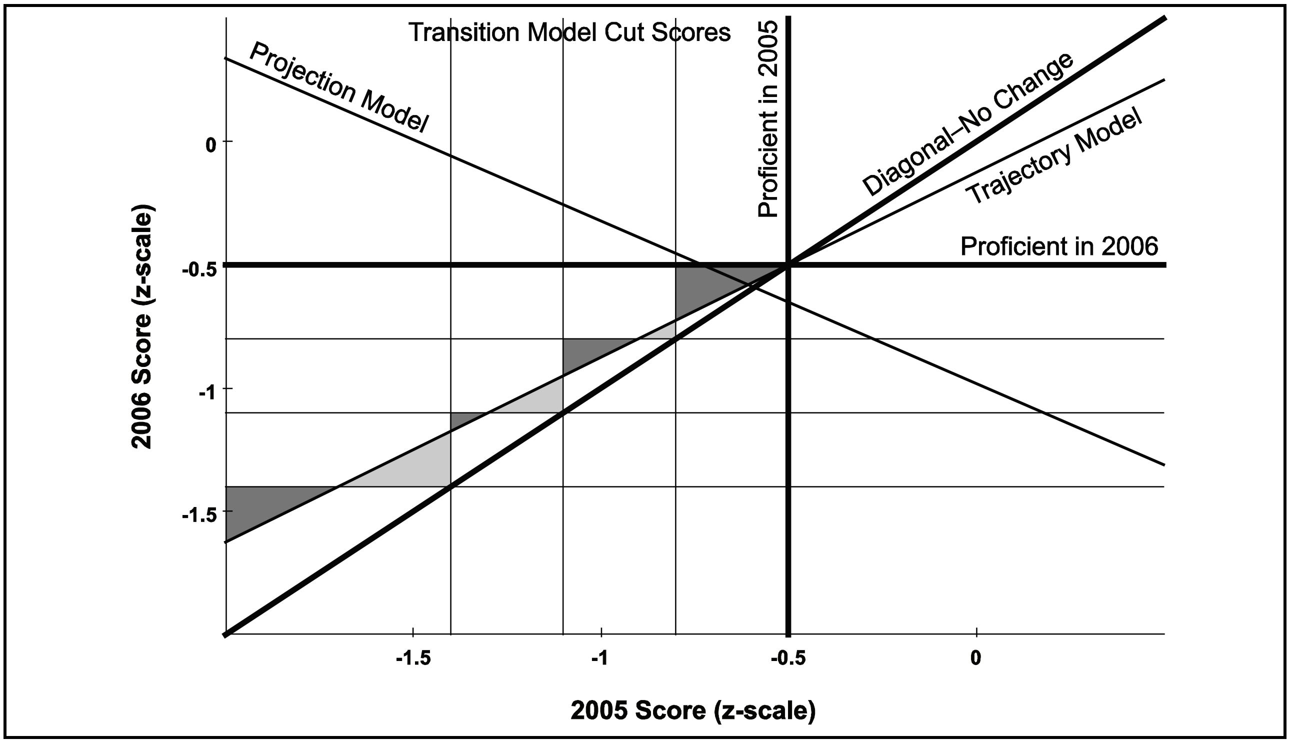


Exhibit reads: The dark-shaded areas correspond to students classified as on-track by the trajectory model but not by the transition matrix model. The light-shaded areas correspond to students classified as on-track by the transition matrix model but not by the trajectory model.

The generic projection model decision rule is illustrated by the line with negative slope labeled "Projection Model" in Exhibit 46. The derivation of this line is explained in Appendix D. Exhibit 47 removes the transition matrix model borders to unclutter the diagram and focuses on the classification area for the projection model. The projection model will classify students who are currently (2006 in this illustration) below proficient as on-track if their scores are in the shaded area in Exhibit 47. The near-perpendicular relationship of the projection model line to the trajectory model line helps to explain the reason for the low percent agreement shown in Exhibit 40.

What is striking about this area is its lack of overlap with the on-track areas for the other two models. The vast majority of students classified as on-track by the projection model are in fact below the main diagonal: they are showing score declines over time, and most were proficient in 2005 but not proficient in 2006. This is a dramatic departure from the logic of transition matrix and trajectory models, but it is not so surprising in the context of regression. When scores have high and similar correlations across years, as is typical of educational test scores, the slopes of projection model lines will be negative (see Appendix D). This translates to the statement, high past scores predict high future scores. Projection models are actually insensitive to the chronological order of grades from which the prediction is made. For example, if a third-grade score is very high, projection models allow that to compensate for a fourth-grade score that is below proficient. This feature of projection models is unchanged by increasing the numbers of years and subjects for prediction; this would increase the accuracy of prediction but not change the nature of the model.

Exhibit 47
The Projection Model Classification Approach

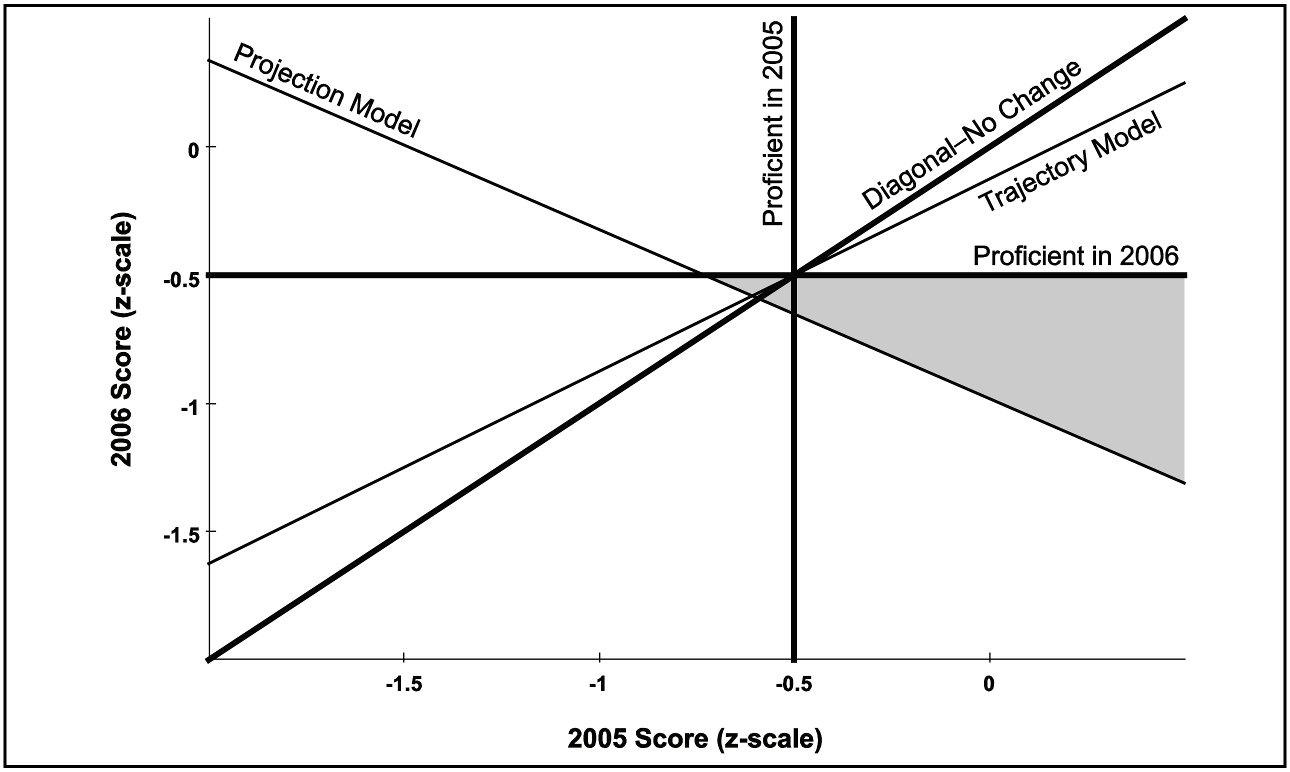


Exhibit reads: The shaded area corresponds to students classified as on-track by the projection model.

Exhibit 47 also allows contextualization of the percent agreement rates in Exhibits 40 and 42. All models agree that students in the tent shaped triangle under the main diagonal and the Projection Model line are not on-track. In this case, however, there is not one single student whom all three models classify as on-track. The projection model area and the trajectory model area only overlap in the small triangle in the center of the figure, bounded by the Projection Model line, the Trajectory Model line, and the Proficient in 2006 line. In the real data example for the Grade 5 Math cohort in 2006, due to the discrete nature of test scores, there are no students in this triangle at all. In this case, all positive projection model classifications are not supported by the other two models, and vice versa.

This stark contrast represents two competing views about the prediction of progress, one educational and one statistical. The educational view is based on a momentum metaphor, ironically both a trajectory and a projection: objects in motion will remain in motion. The statistical view takes the chronological ordering of events into account only as much as proximal correlations are greater than distal correlations and, through regression techniques, offers a statistical prediction that is effective but blind to progress over time.

Exhibit 46 also explains why projection model classification rates are lower than others. Student scores are not uniformly distributed throughout the plot. More students hover close to the diagonal due to the positive correlation of adjacent-grade scores. As a result, more students will be located in the trajectory-area triangle than in the projection-area triangle. This leads to the findings in Exhibit 38. Further consideration of the framework with respect to the underlying bivariate distribution allows for visual explanations of all the bar graphs shown in previous exhibits.

Extending the Framework to the Growth-Only Results

The previous sections applied the generic growth models to a single, common dataset using the status-plus-growth rule as is typically done in the GMPP states. This section extends that analysis by applying the hypothetical "growth-only" rule when employing the various generic growth models to the common dataset. This extension is similar to the application of the "growth-only" rule to the GMPP states in the sections at the beginning of this chapter.

Extensions of the status-plus-growth framework to include use of on-track indictors for all (non-proficient and proficient) students are straightforward and are shown here. Exhibit 48 adds transition matrix model categories above the proficient cut score and shows the growth-only transition matrix model approach. Category maintenance below proficiency is classified as not on-track and category maintenance above proficiency is classified as on-track. The 90 percent agreement rate with the trajectory model is unsurprising here, as the cutoff lines for the transition matrix model zigzag across the cutoff line for the trajectory model.

Exhibit 48
Transition Matrix Model Classification Areas for On-Track Students in 2006

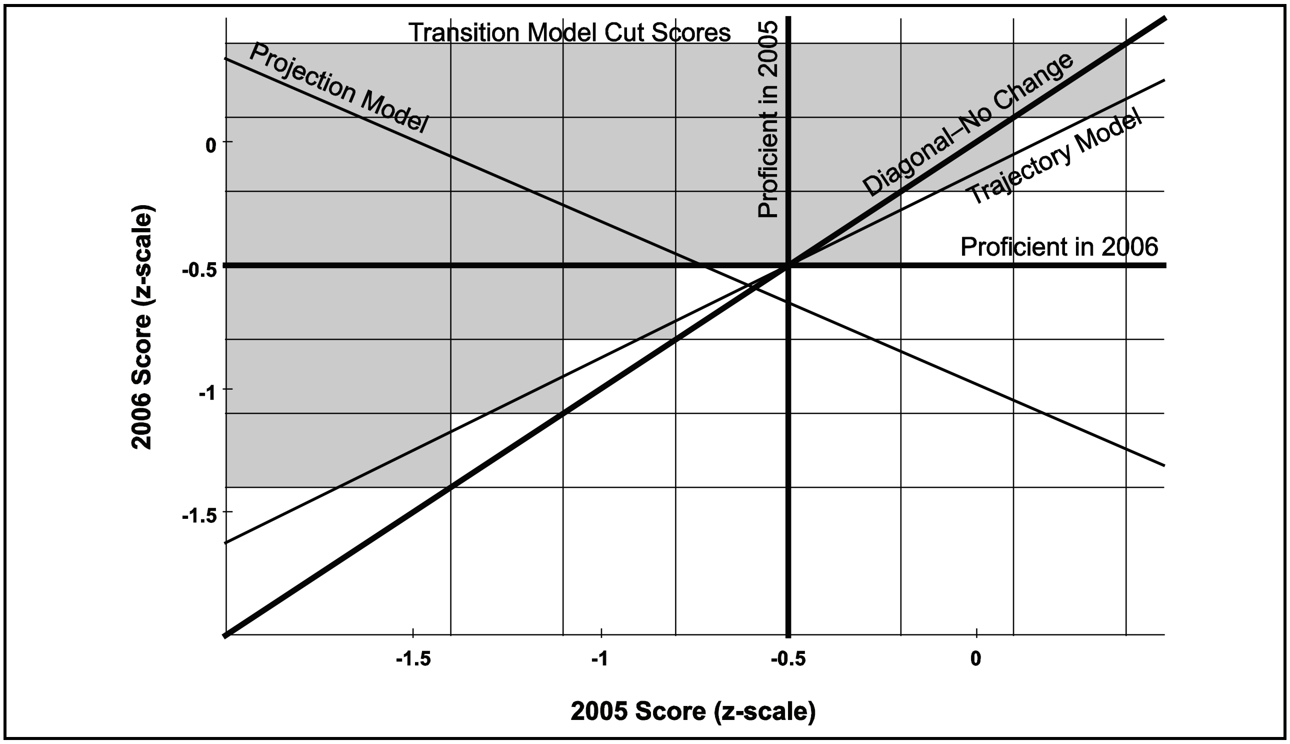
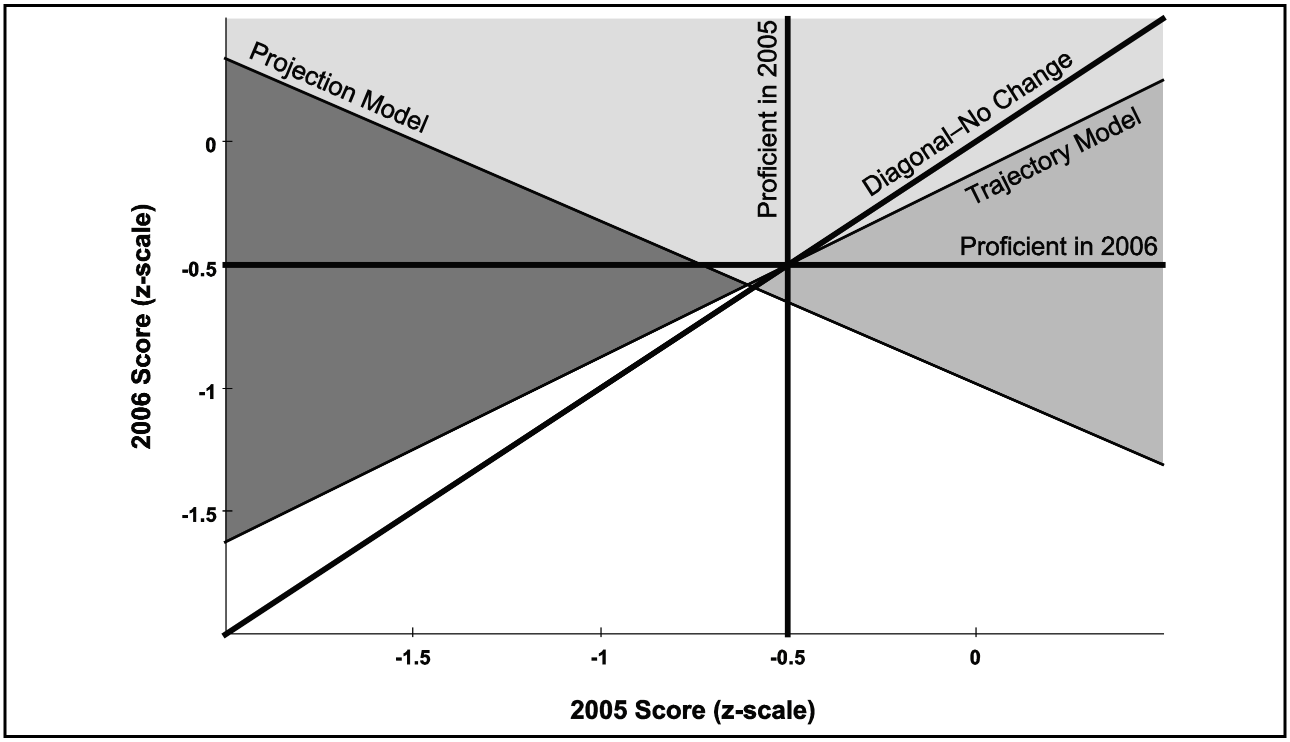


Exhibit 49 shows the contrasting areas of on-track classification for trajectory and projection models under a growth-only approach. The dark-shaded area on the left shows the area where students who score low in the first year but make gains are recognized as on-track by the trajectory model but overlooked by the projection model. The medium-shaded area on the right shows the area where students who score high in the first year but make declines are recognized as on-track by the projection model but passed over by the trajectory model. This figure also makes it clear that decreasing the slope of the Trajectory Model line will increase the agreement between the two models. This is exactly what happens as the time horizon decreases in Grades 6 and 7, and this explains the increasing agreement between trajectory and projection models in those grades.

Exhibit 49
Areas Where Trajectory Models and Projection Models Will Disagree



Finally, Exhibit 49 explains why projection models classify relatively fewer students in status-plus-growth models but relatively more students in growth-only models. Due to the location of the bivariate scatterplot, centered on (0,0) and stretching along the diagonal, the projection model's shaded area (including both the medium- and light-shaded regions) simply captures more students. Importantly, the framework allows for the visualization of the cut-score dependence of this finding. If the cut scores were higher, the trajectory model line would drag up the diagonal, but the projection model line would move straight up (changing only the intercept), thus the projection model and trajectory model classification rates would both decline but become more similar.

School AYP Simulations Based on the Generic Models

The student-level analyses presented above indicate that the models yield different rates of students classified as on-track to proficiency, and that the students so identified are quite different in the projection model than with the transition matrix and trajectory models. The last step of the generic model comparison is to aggregate the student results to the school level in order to simulate AYP determinations.

Two methods of determining AYP are used here. The first compares closely to the method used in all of the pilot states except Delaware, what was referred to as "status-plus-growth" in Chapter I. This involves first calculating the percentage of students in the school who are proficient in both reading and mathematics. If this percentage meets or exceeds the annual measurable objective (AMO), the school made AYP and is classified as having made it by status. If the school did not make AYP by status, then the on-track data from the growth model are used for all non-proficient students. If the addition of the on-track data to the proficient data raises the school percentage to meet or exceed the AMO, the school is classified as having made AYP by growth.

The second method parallels the procedure that Delaware followed, whereby the growth model on-track data are used first. If the percentage on-track exceeds the AMO, the school made AYP by growth. If the percentage on-track was less than the AMO, then students who were proficient but not on-track to maintain proficiency are added to the on-track numbers. If that percentage meets or exceeds the AMO, the school is classified as making AYP by growth-plus-status.

Both methods require setting subgroup minimum size requirements, and we selected 40 students as the minimum. The proficiency cut score level used in this simulation is the -0.50 level used in the initial student-level analysis.

The results for the first (status-plus-growth) method are shown in Exhibit 50. Using the -0.50 cut point for proficiency and the average AMO levels for North Carolina in 2006–07 (56 percent), only about 20 percent of the schools met the AMO and would make AYP with the status model. To determine the number of additional schools that would make AYP using the growth data, the growth model results for non-proficient students in subgroups that did not meet the AMO are then added to the number of proficient students in those subgroups, and the pass rate was recalculated. If all percentages meet the AMO, the school is classified as making AYP by growth. The trajectory model yielded a slightly higher percentage, about 15 percent, than the projection and transition matrix models. These results are consistent with the student level findings reported earlier, that the trajectory model generally identifies a higher proportion of non-proficient students as on-track to proficiency than the other models.

Exhibit 50
Percentage of Schools Making AYP by Status and by Growth Based on Results from the Different Types of Generic Growth Models

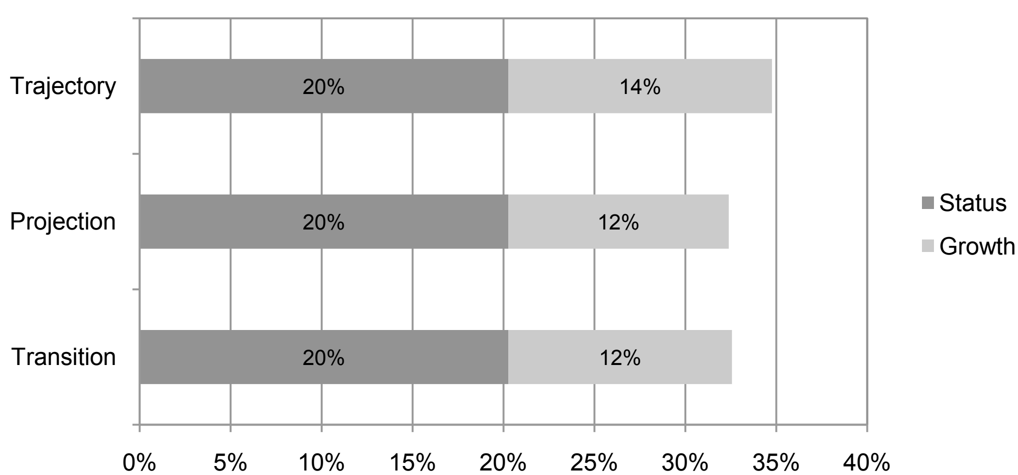


Exhibit reads: In addition to the 20 percent of schools making AYP by status, the trajectory model classifies another 14 percent of schools as making AYP.

The AYP results from the different types of models change markedly if the growth model results are given priority and schools are first evaluated on whether they meet the AMO using growth-only results. Exhibit 51 shows that only 3 percent of the schools made AYP using the trajectory model to classify students as on-track to proficiency and then comparing the percentages of students on-track with the AMO. The transition matrix model yielded similar results but with even fewer schools making AYP by growth only. In contrast, the projection model results in about 30 percent of the schools making AYP by growth only.

This is consistent with the student level results and the theoretical framework laid out earlier, both of which revealed the tendency of the projection model to classify more students as on-track to maintain or exceed proficiency in a growth-only context. As the previous section described, the projection model tends to operationalize student growth more as high average scores. With moderately low (but realistic) cut scores of -0.5, there are many high average scorers. In contrast, trajectory and transition matrix models require growth in a growth-only context, and there are many decliners, whether proficient or not. This discrepancy results in projection models having a relatively large impact in a growth-only setting but a relatively small impact in a status-plus-growth setting at both the student and school levels.

Exhibit 51
Percentage of Schools Making AYP by Growth-Only and Growth-Plus-Status Based on Results from the Different Types of Generic Growth Models.

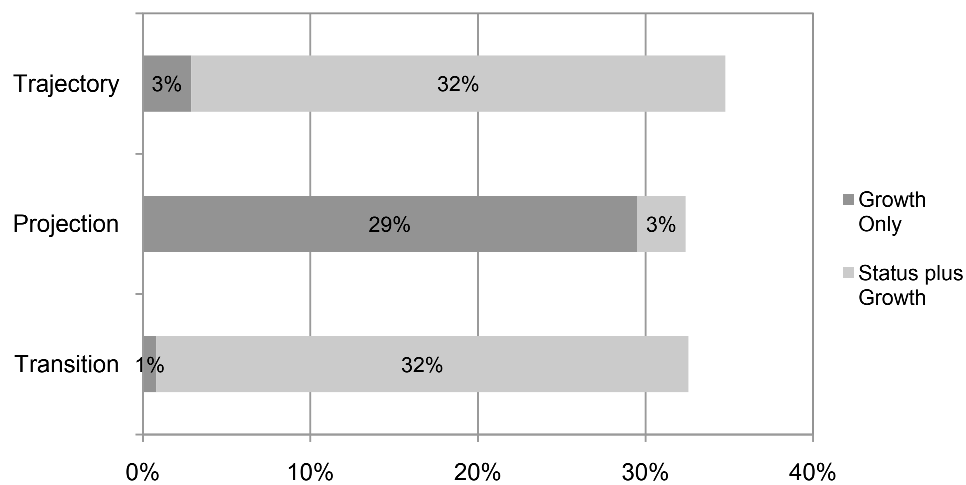


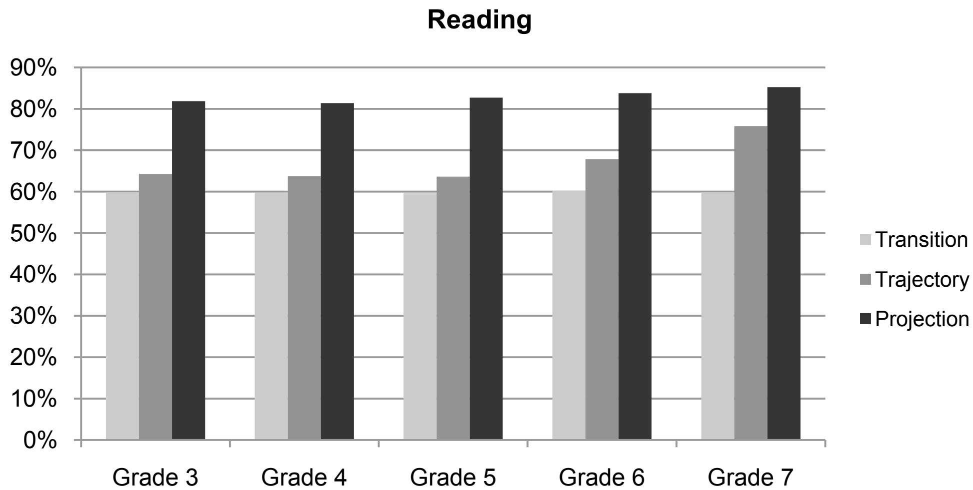
Exhibit reads: The trajectory model classifies 3 percent of schools as making AYP by growth only and an additional 32 percent of schools as making AYP if status criteria are added after growth is applied.

Consistency of AYP Determinations With Future Student Learning

This analysis builds on results of previous comparisons of generic models to assess the predictive accuracy of growth model results at the student level. To compare correct-classification rates, a full replication of all analyses was performed, substituting a 2003 year of focus for the original 2006 year of focus. The numbers of students and the match rates were smaller, but results did not differ substantively from the 2006 results. Unlike the 2006 year of focus, which was the most recent year of representative data availability, the 2003 year of focus allows a later validity check for 2003 classifications. Grades 3–5 on-track predictions could be evaluated using the 2006 data, three years later, and Grade 6 and 7 predictions could be evaluated two years and one year later in 2005 and 2004 respectively.

Exhibit 52 shows the correct-classification rates by model and grade for a growth-only model: how often on-track students were actually proficient and not-on-track students were actually non-proficient. As expected of a regression model, the projection model correct-classification rate is higher, by around 20 percentage points over the transition matrix and trajectory models in the lower grades. As the time horizon shrinks in Grades 6 and 7, and the trajectory model classification line more closely approaches the projection model's classification line, the projection model advantage drops under 10 percentage points.

Exhibit 52
Correct Classification Rates by Model and Grade, Reading and Mathematics



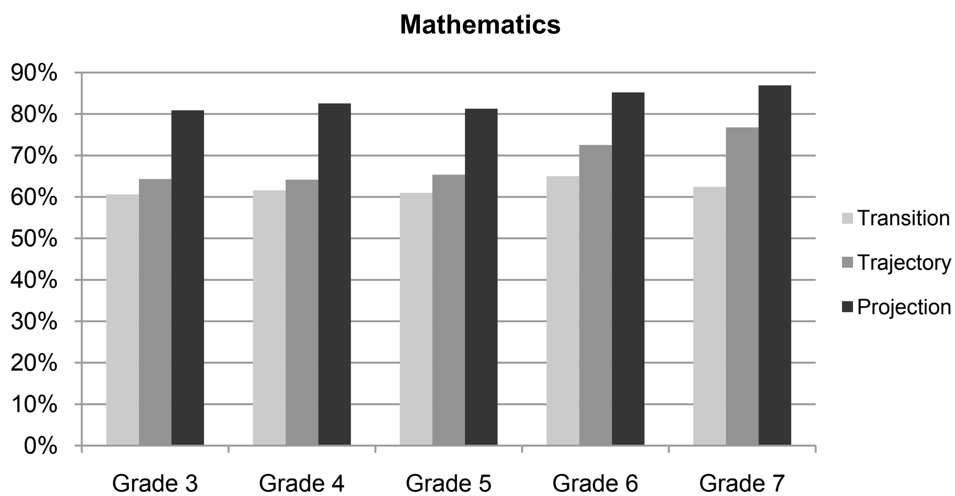
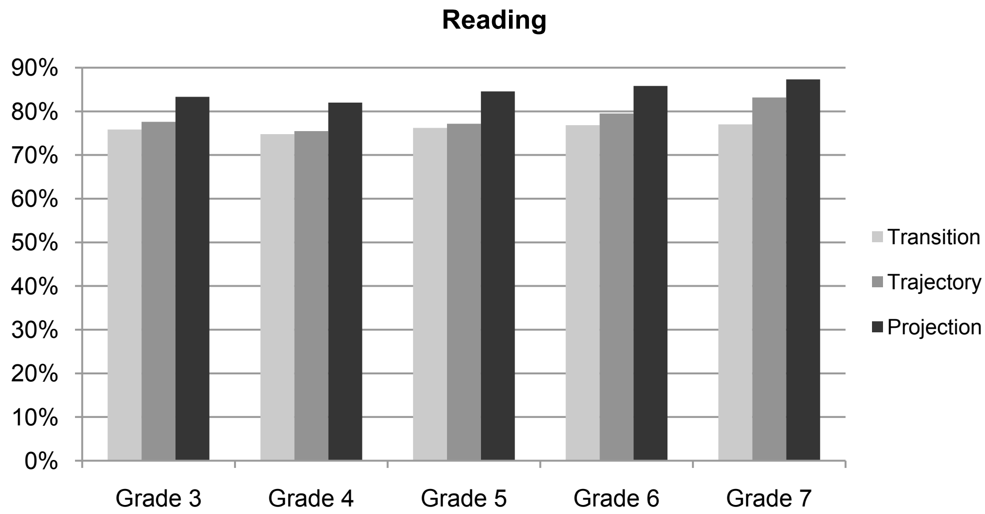


Exhibit reads: Of the third-graders who were classified with the generic transition model as on-track or off-track in reading, 59.9 percent turned out to achieve their predicted status of proficiency or non-proficiency.

Whereas Exhibit 52 shows correct classifications for both on-track and non-on-track students, Exhibit 53 reports the percentage of on-track-designated students that actually reach proficiency. Naturally, these percentages are higher, with the projection model's advantage decreasing from 7 percentage points in the lower grades to less than 5 percentage points in the higher grades. The ranking of the models is the same across all subjects and grades.

Exhibit 53
Percentage of On-Track-Classified Students Who Achieve Proficiency by Model and Grade, Reading and Mathematics



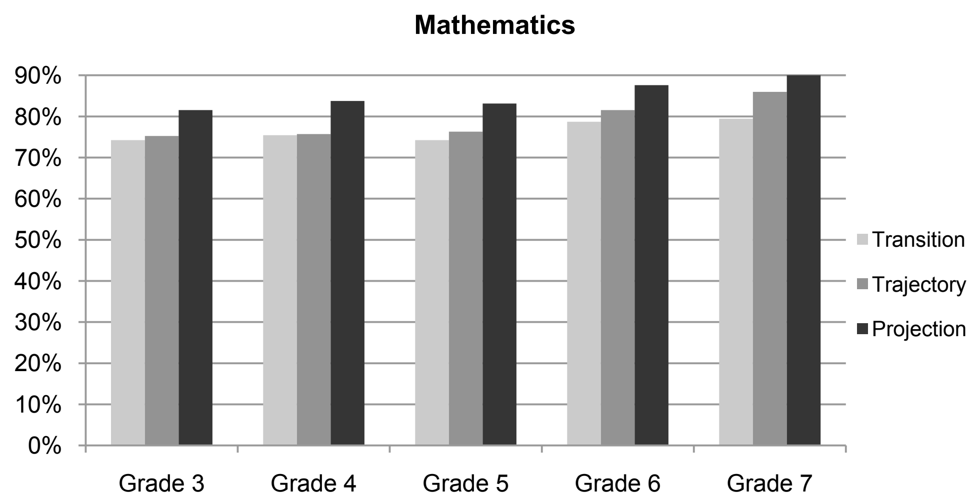


Exhibit reads: In reading, 75.8 percent of the Grade 3 students that the generic transition model classified as on-track turned out to achieve proficiency.

However, trajectory models may have utility through appealing to the logic of teachers and students in schools. Students identified by the trajectory model may indeed make a gain due to measurement error, but it may make more educational sense to teachers and administrators to reward observed gains than high-averaging declines. These models thus contrast both in setting incentives and describing the impact of policies. These models establish both predictions and examples for the kinds of students that deserve recognition. Given these tradeoffs, states may choose their balance of priorities as they develop future growth models.

Comparison of Generic Growth Models: Summary and Conclusions

The projection model functions in stark contrast with transition matrix and trajectory models in terms of how they classify students as on-track and not no-track. When a projection model classifies a non-proficient student as on-track, the probability that a transition matrix or trajectory model agrees is near zero, and vice versa. Agreement rates for non-proficient students are around 60 percent due to agreement about students not on-track.

For status-plus-growth models (in which only non-proficient students are classified as on-track or not on-track), projection models will have the least impact, affecting only 10–20 percent of eligible (non-proficient) students while transition matrix and trajectory models affect over 20 percent. States with higher cut scores will have heightened differences between the model types and a lower proportional impact of projection models on eligible students.

For growth-only models, in which all students are classified either as on-track or not on-track, projection models will be the most lenient, classifying around 70 percent of students as on-track for typical cut scores, around 10 percentage points greater than the other models.

A graphical framework allows for visualization of the differences between the models. It also affords clear predictions under alternative assumptions about distributions and cut scores.

Using the data from the generic models to simulate school AYP determinations, the models do not yield large differences in the percentages of schools making AYP for status-plus-growth models. The models differ much more under a growth-only approach, in which very few schools would make AYP with a trajectory or transition matrix model, and many schools would make AYP with a projection model.

The underlying reasons for the contrast between the projection model and the transition matrix or trajectory models relate to differing objectives guiding the models. Projection models are designed to maximize the accuracy of a prediction about whether a student will meet or exceed future proficiency standards. From a statistical standpoint, the best predictor of future achievement is past achievement. As a result, relatively few students with records of low achievement are predicted to meet or exceed future proficiency standards, while students with records of high achievement are very likely to be predicted to do so. Transition matrix and trajectory models, in contrast, are designed to identify specific growth targets that each student must attain in order to meet or exceed the future proficiency standards, given his or her benchmark score (usually from the last test administration). Reflecting these different guiding objectives, projection models classify as on-track a large number of previously proficient decliners (many of them correctly), whereas transition matrix and trajectory models classify as on-track a large number of lower-scoring gainers (some of them incorrectly). The educational significance of this contrast is briefly discussed.

Effects of Different Growth Standards for School Accountability

The previous sections compared three types of growth models. According to the GMPP policy, each of these follows the bright-line principle of ensuring universal proficiency by 2014. Thus, these models set individual growth standards quantifying adequate progress to proficiency by a time horizon, graduation, or 2014: These are growth-to-proficiency models. Alternative models exist that do not depend or depend less on proficiency standards and time horizons. These models may define adequate growth for each grade or different starting proficiencies, or a common growth standard that can be applied to all students. It is important to note that such alternatives are not currently allowed for use in *ESEA*-mandated AYP reporting. Nonetheless, it is potentially useful to make controlled comparisons between growth-to-proficiency models and alternative types of growth models.

The analysis that follows assesses the hypothetical impact on school AYP determinations of using a different standard of "adequate yearly growth" than the growth-to-proficiency piloted under the GMPP. Toward this end, we constructed a growth indicator based on the difference between the proficiency cut points of successive grade levels. Non-proficient students who gain more than that difference are classified as on-track, even though their growth trajectory may not be steep enough to reach proficiency within the three- or four-year time frame of the growth model. For this analysis, the data are drawn from the actual Florida database (along with the actual Florida proficiency cut scores) and include students with test scores in 2007–08 and 2006–07. Florida's data are well suited for this simulation because the state provided longitudinally linked scale scores for their students and the scale scores were vertically equated.

A comparison of results from the actual GMPP growth-to-proficiency model with the alternative growth standard is shown in Exhibit 54. The upper panel shows the actual GMPP and the lower panel shows the alternative growth standard results. The first vertical bar for each grade level in both panels is the percentage of students that were proficient or higher in both reading and mathematics at each grade level; these ranged from 48 to 62 percent across these grade levels. The second bar (labeled "GMPP on track" in the upper panel and "alt. growth" in the lower) is the percentage of students who were classified as making "adequate growth" under the respective growth models. The percentages of students who were on-track per the GMPP growth model (upper panel) ranged from about 44 to 52 percent across grades. In contrast, the percentages that met the alternative growth standard (i.e., that gained an amount equal to or greater than the difference between their current grade level proficiency cut score and the proficiency cut score for the prior grade level) were much lower, ranging from 16 to 23 percent except for grade 7 where 38 percent of students were proficient under the alternative growth standard. The large increase in grade 7 was likely a reflection of an unusually small distance between the proficiency cut scores for grade 6 and grade 7.

Exhibit 54
Percentages of Students Scoring At or Above Proficiency, Making GMPP Annual Growth Targets (Upper Panel), Making Alternative Annual Growth Standard (Lower Panel), and Meeting Either Proficiency or Growth Target, Florida Data from 2006–07 and 2007–08

(Upper Panel: GMPP Growth)

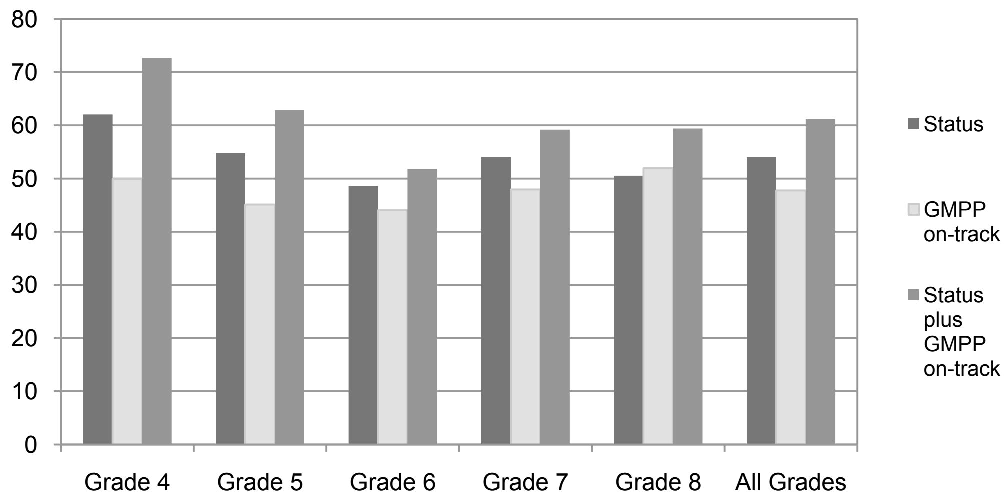


Exhibit reads: In 2007–08, 62 percent of the grade 4 students in Florida scored at or above the proficiency cut scores in both reading and mathematics. Fifty percent of the students were classified as on-track by the Florida GMPP growth model, and 73 percent were either proficient or on-track.

(Lower Panel: Alternative Growth)

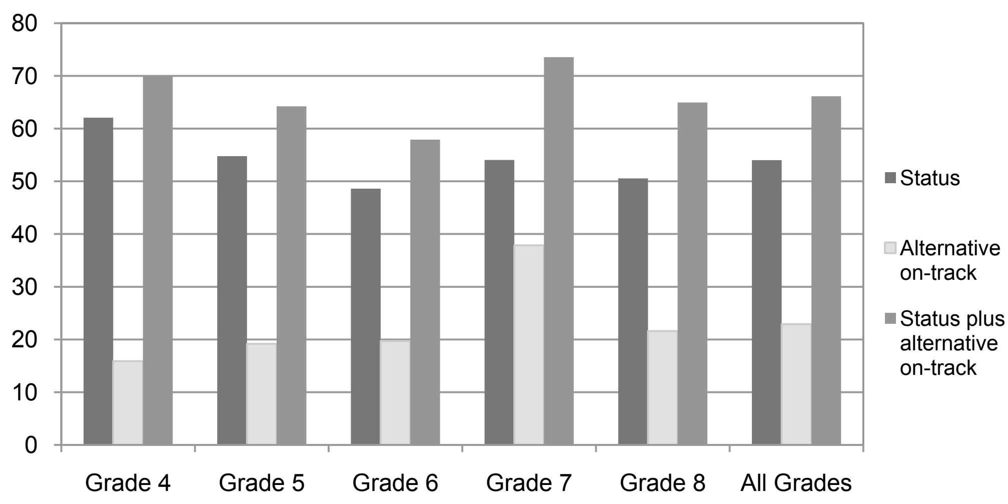


Exhibit reads: In 2007–08, 62 percent of the grade 4 students in Florida scored at or above the proficiency cut scores in both reading and mathematics. Sixteen percent of the students were classified as making adequate annual growth by the hypothetical alternative growth model, and 70 percent were either proficient or making adequate growth.

This simulation shows that the percentages of students meeting this alternative standard of growth are lower than the percentage of proficient students in both reading and mathematics.

However, adding non-proficient students who met the growth standard to the pool of proficient students would increase the overall rates of students who are arguably performing adequately (i.e., proficient or making reasonable progress over the past year).

A possible shortcoming of this method of assessing adequate yearly growth is that it may be negatively correlated with initial achievement level, such that students with higher initial achievement gain less and are more likely to not make adequate yearly growth by this standard. This could occur because the state tests are designed to measure achievement with respect to grade level standards and may be prone to ceiling effects. If the model is less sensitive to gains made by higher-achieving students, the minimally acceptable growth increment could be conditioned on initial level of achievement.

Exhibit 55 shows evidence consistent with ceiling effects in reading and mathematics in Florida. This graph plots the average scale-score gains from 2006–07 to 2007–08 on the vertical scale (y-axis) by the decile of the student's 2006–07 score. The average gains in both subjects are greatest among students with the lowest scores in 2006–07. The average gains gradually decline from decile 2 to decile 8 and then drop more markedly in decile 10. These results suggest that, at least for this assessment system, somewhat different approaches to assessing adequate yearly growth may be required if a growth-only model is being considered. In a status-plus-growth model, however, lower expected growth is unlikely to have an impact on how high-achieving students are classified, because these students are proficient and the growth model is only applied to non-proficient students.

Exhibit 55
Average One-Year Gains in Mathematics and Reading, by Base Year Achievement Decile

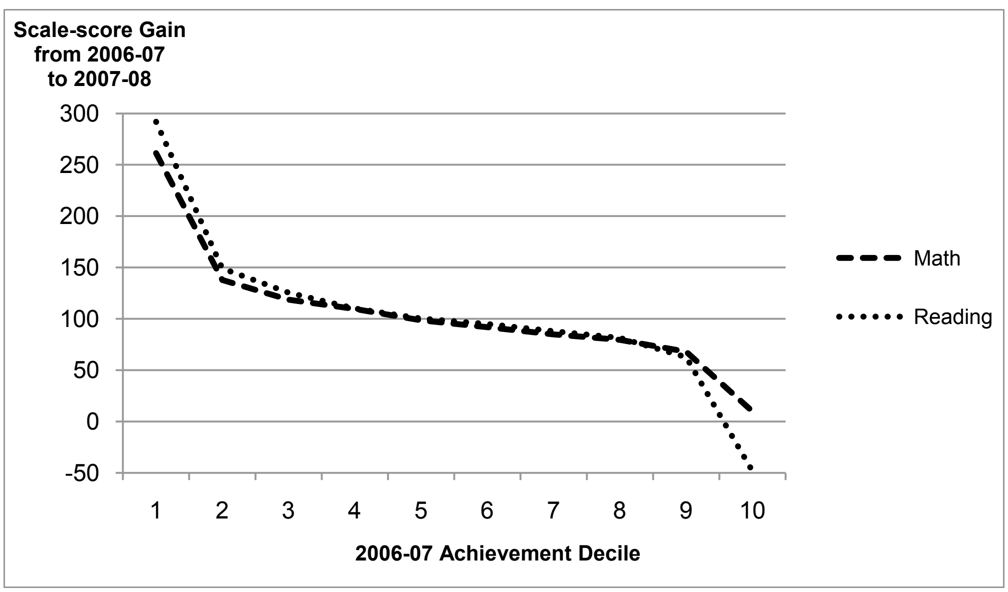


Exhibit reads: Florida students who scored in the lowest decile had average one-year gains of 260 scale-score points in mathematics and 300 scale-score points in reading.

This alternative growth standard is similar to some state (as opposed to federal) accountability systems in that it is not grounded in students being on-track to meet or exceed proficiency cut scores. However, this hypothetical alternative probably yields lower rates of students and schools classified as making adequate growth than the growth models used in most state accountability systems. State accountability growth models often define adequate growth as “any growth” while this hypothetical formulation defines adequate growth in terms of a specific common yardstick for each grade.

Effects of Longitudinal Matching

The final aspect of growth models examined here is the extent to which the state assessment systems were able to calculate growth indicators for their students. Growth indicators are generally based on at least two years of achievement data. However, some students in each school year will not have prior test score data because they transferred in from out of state or from a private school, were absent on the test dates, or were excluded from the prior year testing for one reason or another. The percentage of students with growth indicators is referred to as the match rate (all of the pilot states except Ohio and Tennessee required two consecutive years of test scores in order to calculate growth indicators).

ESEA requires that schools test each year at least 95 percent of their students in both mathematics and reading or language arts. However, it is possible for the percentage of students with test scores from both the current and the prior year (students whose scores can be ‘matched’) to be lower than 95 percent.

The match rate has potentially important implications for the validity of a growth model: if a growth-only model is applied to all students in an *ESEA* reporting group in order to assess whether the AMO was met, then only the matched students are used to determine proficiency rates. Among the nine states examined here, only Delaware, Florida, Ohio, and Tennessee applied their growth models to all students in *ESEA* reporting groups. In Delaware, growth model results were considered first in making AYP decisions for each *ESEA* reporting group. Florida, Ohio, and Tennessee applied growth results after status and safe-harbor but used the growth model results for *all students* within *ESEA* reporting groups that did not make AYP by status or safe-harbor. The other states, in contrast, only used growth results for non-proficient students within reporting groups that did not make AYP by growth or safe-harbor. In those states, the non-proficient students without matched data would continue to be counted as non-proficient and would not be subtracted from the denominator for their respective reporting group’s AYP calculations.

In order to gain some information on how the results of growth models like the ones piloted in Delaware, Florida, Ohio, and Tennessee might be affected by match rates, matched and non-matched students were compared in seven of the states (Iowa and Tennessee did not provide sufficient information to calculate match rates). Exhibit 56 presents the match rates and average test score differences among matched and non-matched students in the states that provided scale-score data to the evaluation project. The match rates are on the number of students who have the two years (2006–07 and 2007–08 school years) of test scores needed to calculate growth. The number of students with two years of data was divided by the number of students in the indicated grade levels who had either or both a reading and a mathematics test score in the 2007–08 school year. The seven states generally had match rates of 91 percent or higher based on the supplied student data (Exhibit 56).

Exhibit 56
Two-Year Match Rates for Students and Differences Between Matched and Unmatched Grade-Standardized Student Scores for Reading and Math, by State, 2007–08

Pilot States[a]	Grades	Number of Students w/ Test Scores from 2007–08	Percent with Test Scores from 2006–07	Difference (SDs) In Reading	Difference (SDs) in Math
Alaska	4–8	46,692	92%	0.08	0.14
Arizona	4–7	286,455	95%	0.39	0.37
Arkansas	5–7	106,133	91%	0.26	0.27
Delaware	4–8	43,332	93%	0.19	0.19
Florida	4–8	965,325	91%	0.37	0.40
North Carolina	3–8	663,504	93%	0.48	0.58
Ohio	4–8	627,208	93%	1.65	1.56

Exhibit reads: Of the 46,692 students in grades four through eight in Alaska who had either a reading or mathematics test score in 2007–08, 92 percent had matched scores. The mean achievement difference in scale scores between those who were matched and those who were unmatched was 0.08 standard deviations in reading, and 0.14 standard deviations in mathematics.

[a] Iowa and Tennessee did not provide sufficient information to calculate match rates.

Source: The Alaska, Arizona, Arkansas, Delaware, Florida, North Carolina, and Ohio state departments of education

In these seven states, reading and mathematics scale scores were available and the mean scores of matched and unmatched students could be compared. Unmatched students had substantially lower average test scores than matched students. Taken together, these results indicate that while the match rates were high, the matched population may have significantly higher levels of achievement than the unmatched group and thus be more likely to be proficient. The implications are less clear for the proportion of students identified under the growth models to be on-track to proficiency. As seen in the section on the effects of different types of models, lower-performing students are more likely to be classified as on-track under trajectory and transition matrix models than under projection models. To that extent, reliance on growth-only data for AYP determinations may increase the likelihood a school will reach its AMO under a projection model but may not have that effect under the other models. While the match rates in the pilot states that provided the information were generally very high, the potential for a few exclusions to affect AYP results suggests that further study is needed to determine the validity of using non-matched status results rather than excluding them from the growth-only applications. In any case, these points underscore the importance of maintaining high match rates if growth-only outcomes are applied to accountability decisions.

Conclusions

This chapter addressed a number of hypothetical questions about the pilot growth models. The first question addressed was the extent to which the results of using only the growth models' on-track-to-proficiency results would diverge from the status model results among the schools that made AYP by the status model in 2007–08. This analysis identified the schools currently performing sufficiently well to make AYP by status criteria but, according to the state's growth model, not making sufficient annual gains to reach the AMO by growth-only criteria. These schools would then be expected not to continue to make AYP if their pattern of low growth continued. For all nine states combined, 62 percent of the schools that made AYP by status also met their reading and mathematics AMOs using just the growth criteria. The states varied considerably, ranging from only 46 percent in Arizona and 47 percent in Arkansas and North Carolina to 75 percent or more in Ohio, Delaware, and Tennessee.

The percentages of schools that made AYP by safe-harbor and that also met or exceeded their AMO under the growth-only criteria were much lower (28 percent overall). The percentages did not exceed 30 percent in any states except Arkansas (64 percent) and Ohio (45 percent).

The largest share of this chapter consisted of an analysis of the effects of different types of growth models on student on-track and school AYP determinations. The projection model functions in stark contrast with transition matrix and trajectory models. When a projection model classifies a non-proficient student as on-track, the probability that a transition matrix or trajectory model agrees is near zero, and vice versa. Agreement rates for non-proficient students are around 60 percent due to agreement about students not on track.

For status-plus-growth models, projection models will have the least impact, affecting only 10–20 percent of eligible (non-proficient) students while transition matrix and trajectory models affect over 20 percent. States with higher cut scores will have heightened differences between the model types and a lower proportional impact of projection models on eligible students. In contrast, for growth-only models, projection models will classify the greatest numbers of students as on-track.

The models do not yield large differences in the percentages of schools making AYP when non-proficient students who are on-track to proficiency according to each type of model are added to the numbers of students meeting or exceeding the proficiency cut point. The models differ much more when a growth-only calculation is made first, followed by growth-plus-status calculations. Under the growth-only simulation using realistic cut scores and standard AMOs, very few schools would make AYP with a trajectory or transition matrix model while most schools would make AYP with a projection model.

The student-level results from the generic model comparisons were used to compare the models in terms of their predictive accuracy of on-track students actually attaining proficiency. Projection models have the highest correct classification rates for future proficiency: over 80 percent. These rates are 5 to 20 percentage points higher than trajectory and transition matrix models, depending on the grade level and proximity to the growth model time horizon.

The transition matrix model acts as a coarse, categorical approximation of the trajectory model, with agreement rates over 90 percent. The overlap is greatest when time horizons for the trajectory model are long or when the number of categories for the transition matrix model is large.

An alternative standard of annual growth to the *ESEA*-mandated growth-to-proficiency models is the difference between the proficiency cut scores in successive grade levels. While not acceptable under current regulations for determining school AYP, students gaining that amount or more on a vertically scaled assessment would be considered to make "adequate yearly growth" regardless of whether they are proficient or on-track to become proficient. Adding the non-proficient students who met this alternative growth standard to the pool of proficient students would increase the overall rates of students who are arguably performing adequately (i.e., proficient or making reasonable progress over the past year).

The final aspect of growth models examined here was match rates. In this chapter we estimated the extent to which the pilot states had reading and mathematics test scores from both 2007–08 and 2006–07 available, and whether the 2007–08 test scores differed for the students with and without the 2006–07 scores. The results indicate that while the match rates were high (91 percent or greater in the seven states providing the necessary data), the matched population may have significantly higher levels of achievement than the unmatched group and thus be more likely to be proficient.

Conclusions

This report was designed to answer two main questions about the implementation of the Growth Model Pilot Project (GMPP) under the *Elementary and Secondary Education Act*, as amended by the *No Child Left Behind Act of 2001*. The two questions and brief summaries of the study results based on analyses of data provided by nine pilot grantee states from 2006–07 and 2007–08 that are the subject of this evaluation (Alaska, Arizona, Arkansas, Delaware, Florida, Iowa, North Carolina, Ohio (for 2007–08 only), and Tennessee) are as follows:

How have states in the pilot project implemented growth models?

While the models approved for the nine pilot grantee states that are the subject of this evaluation differ from one another in a number of important ways, all use state-specific assessment data to measure student progress and proficiency, and the method of incorporating growth outcomes in adequate yearly progress (AYP) determinations was generally the same. Eight of the nine pilot states applied growth criteria only *after* schools have failed to make AYP under the status and safe-harbor provisions, rather than determining AYP solely on the basis of student improvement—a "status plus growth" model. Delaware, in contrast, applied growth criteria first, and then applied the status model and safe-harbor provisions, respectively, to schools that did not make AYP under the growth model.

When a school is designated as "making AYP under growth," this means that use of the growth model changed the designation for one or more targeted *ESEA* subgroups. Within the affected subgroup(s) for a given school, the growth criterion is usually applied only to the students who did not achieve at or above the proficiency level. Simply stated, "making AYP under growth," as defined by the GMPP, does not mean that all students are on-track to proficiency, and it can even mean only one non-proficient student is on-track if a sufficient number of others in the subgroup are proficient. Eight of the nine pilot states used the status-plus-growth model (Alaska, Arizona, Arkansas, Florida, Iowa, North Carolina, Ohio, and Tennessee). Delaware used a growth-plus-status model for determining AYP.

The models approved for the pilot study vary in how they (1) establish growth expectations for students, and (2) determine whether individual students are "on-track" to reach proficiency in the allotted time frame. We have identified three basic types of growth models being used in the GMPP: the *transition matrix* model (which evaluates student progress from year to year in terms of a relatively small set of discrete performance levels), the *trajectory* model (which uses the gap between a baseline test score and a performance standard several years out to calculate the amount of growth required to become proficient), and the *projection* model (which uses current and past test scores to statistically predict performance several years ahead, using new test scores to update projections of student performance).

States using a given type of model varied greatly in the extent to which schools were identified as making AYP by growth within the GMPP framework. Despite the lack of consistency from state to state, the choice of model type can be consequential, as demonstrated in Chapter IV with the side-by-side comparisons of generic versions of the three types of models using a common dataset. However, the variability in state-by-state AYP results is consistent with the latitude that states had in how they defined proficiency levels and implemented (or not) the various provisions within *ESEA* for recognizing AYP under the status model including confidence intervals and multiyear averaging. Those provisions (and safe-harbor) are generally designed to reduce the chance of incorrectly identifying schools as not making AYP, and the growth models were primarily adopted to further reduce that chance. But if the status provisions were already capturing most of the schools that might have been identified as making AYP by growth, then the numbers of schools identified as growth schools by the GMPP would be correspondingly reduced.

How did each pilot state's growth model affect the number and kinds of schools that make AYP?

The nine pilot states in 2007–08 that are the main subject of this evaluation provided both school-level (ED*Facts*) and the student-level data to the evaluation project and these data were used to address five sub-questions.

How many schools made AYP under the growth model that would not have made it under the* ESEA *status model?

The designs of the growth models in these states included only those students who (a) did not reach proficiency levels in reading or language arts and mathematics, and (b) were members of *ESEA* reporting groups that did not reach their Annual Measurable Objectives (AMOs) or obtain AYP via safe-harbor provisions. The pilot models simply added growth criteria to the traditional status plus safe-harbor model for determining AYP; thus growth criteria could only increase the number of schools making AYP. The number of schools identified as making AYP by growth under the GMPP in 2007–08 ranged from 2 percent or fewer of all schools in Alaska, Arizona, Iowa, North Carolina, and Tennessee, to 3 percent in Delaware, 5 percent in Florida, 6 percent in Arkansas, and 34 percent of all schools in Ohio. The schools making AYP uniquely by growth represented percentage increases in the numbers of schools making AYP that ranged from highs of 102 percent in Ohio, 24 percent in Florida, and 10 percent in Arkansas, to lows of 4 percent or fewer in Alaska, Arizona, Delaware, Iowa, North Carolina, and Tennessee. Expressed as percentages of the schools that did not make AYP under status or safe-harbor, the impact of the GMPP on identifying additional schools as making AYP ranged as high as 50 percent in Ohio, 13 percent in Arkansas, and 10 percent in Tennessee.

Are AYP outcomes under the growth models related to school demographics and organizational characteristics?

Schools enrolling higher proportions of low-income and minority students were more likely to make AYP under growth in the status-plus-growth framework than were schools enrolling higher proportions of more affluent and nonminority students.

How many schools that made AYP under the* ESEA *status model would also have made it if the growth criteria were used exclusively for assessing whether schools met their AMOs?

This question was addressed by drawing on the student data provided by the pilot states to calculate whether each school met its AMOs for reading and mathematics using the on-track to proficiency indicators for each student. For all nine states combined, 67 percent of the schools that made AYP by status also met their reading and mathematics AMOs using just the growth criteria. Results varied widely among the states: only 46 percent of the Arizona schools that made AYP by status also met their AMO using growth criteria alone, but almost all of the Ohio (99 percent) and 82 percent of the Tennessee schools that made AYP by status also met their AMO with the growth criteria.

What is the relationship between AYP status under the growth model and* ESEA *safe-harbor provisions?

Eight of the nine pilot states had at least one school classified in ED*Facts* as having made AYP by safe-harbor provisions. The percentages of schools that made AYP by safe-harbor and that also met or exceeded their AMO under the growth-only criteria were much lower (38 percent overall) than schools that made both status and growth-only and were inconsistent across the eight states with safe-harbor schools, ranging from 4 percent in Alaska to 90 percent in Ohio.

How do the main types of models compare in terms of the number of students they identify as on-track and the number of schools that meet their AMO by growth criteria?

The projection model functions in stark contrast with transition matrix and trajectory models. When a projection model classifies a non-proficient student as on-track, the probability that a transition matrix or trajectory model agrees is near zero, and vice versa. Overall agreement rates for non-proficient students are around 60 percent due to agreement about students not on track.

The transition matrix model acts as a coarse, categorical approximation of the trajectory model, with agreement rates over 90 percent. The overlap is greatest when time horizons for the trajectory model are long or when the number of categories for the transition matrix model is large.

For growth-plus-status models applied to a common dataset, projection models will have the least impact, affecting only 10–20 percent of eligible (non-proficient) students while transition matrix and trajectory models affect over 20 percent. States with higher cut scores will have heightened differences between the model types and a lower proportional impact of projection models on eligible students. In contrast, for growth-only models, projection models will classify the greatest numbers of students as on-track.

The models do not yield large differences in the percentages of schools making AYP when non-proficient students who are on-track to proficiency according to each type of model are added to the numbers of students meeting or exceeding the proficiency cut point. The models differ much more when a growth-only calculation is used. Under the growth-only simulation using realistic proficiency cut scores and standard AMOs, very few schools would make AYP with a trajectory or transition matrix model while nearly 30 percent of schools would make AYP with a projection model.

How do the main types of models compare in terms of their predictive accuracy of on-track students actually attaining proficiency?

When models are applied to a common dataset, projection models were shown to have the highest correct classification rates for future proficiency: over 80 percent. These rates are 5 to 20 percentage points higher than trajectory and transition matrix models, depending on the grade level and proximity to the growth model time limit.

How would AYP results change if the growth standards were not tied to attainment of proficiency within the growth models' time horizons?

Designed to assess progress toward proficiency within a four-year-or-less time frame, the pilot growth models can require very large annual growth increments for students who start at low levels of achievement. If *ESEA* regulations were changed, other kinds of growth models than the currently required growth-to-proficiency models could be used for AYP determinations. One alternative standard of annual growth that could be used with vertical test score scales is the difference between the proficiency cut scores in successive grade levels. Students gaining that amount or more would be considered to make "adequate yearly growth" regardless of whether they are proficient or on-track to become proficient. A simulation shows that the percentages of students meeting that alternative standard of growth are lower than the percentage of proficient students in both reading and mathematics, but that adding the non-proficient students who met the growth standard to the pool of proficient students would increase the overall rates of students who are arguably performing adequately (i.e., proficient or making reasonable progress over the past year).

To what extent are longitudinal student data required by the growth models available?

We estimated the extent to which the pilot states had reading and mathematics test scores from both 2007–08 and 2006–07 available, and whether the 2007–08 test scores differed for the students with and without the 2006–07 scores. While the match rates were high (90 percent or greater in all states), the matched population may have significantly higher levels of achievement than the unmatched group and thus be more likely to be proficient or on-track to proficiency. To that extent, reliance on growth data with their requirement for matched longitudinal records for AYP determinations may increase the likelihood a school will reach its AMO.

Implications for Future Policy

Implications for future policy relate to reporting of growth model results, selection of growth models, and use of growth models not tied to proficiency standards.

Reporting of growth model results

As used for school AYP purposes in all states except Delaware, the growth information for students was used only for reporting groups that did not make AYP under status criteria or safe-harbor provisions. Consequently, it was generally not possible to determine from the official ED*Facts* reports the extent to which a school designated as making AYP by growth was actually realizing growth among its students. Delaware applied growth criteria before status and safe-harbor criteria and all schools identified as making AYP by growth actually met their AMOs on the basis of Delaware's growth model. An implication for future policy is that other states could clarify their schools' progress by applying growth criteria before status and safe-harbor. Two general ways of doing that can be identified and each has advantages and disadvantages. Both of these options are methods for *reporting* schools' results related to AYP determinations and do not suggest changes that would actually affect whether schools are determined to make AYP. As described here, neither of the methods would affect a school's AYP status. However, the second option is recommended because it fully establishes and reports both the growth and the status information.

Reporting option #1: Applying growth criteria before status and safe-harbor criteria to all schools' AYP determinations

A different possible ordering would involve first using the growth results to assess whether the reporting group reached the AMO, then, if not, using the union of growth and status results to estimate the percentage of students on-track or proficient. Safe-harbor might only then apply to reporting groups that did not reach the AMO through either of these calculations.

The main advantage of applying growth before status and safe-harbor is that it would identify schools that are realizing adequate progress toward universal proficiency. This would clearly distinguish those schools from schools making AYP under status criteria but not realizing growth sufficient to continue meeting their AMOs. The latter are probably not a large number but identifying such schools would serve as an early warning mechanism of possible problems. The exploratory analyses in this report also indicate that applying growth criteria before safe-harbor could usefully reclassify most of the current safe-harbor schools as making AYP by growth and would clearly identify the minority that are not on-track to proficiency and thus headed for sanctions in the near future.

One disadvantage is that schools that are making AYP both by status and growth (arguably the strongest schools) would not be uniquely identified. Another possible disadvantage of applying growth first follows from the fact that the AYP-by-growth determination would be based on students with at least two consecutive years of test data. Matched students are likely to have higher test scores and thus make it more likely that their subgroups and school reach the AMO. While the match rates in the pilot states that provided the information were generally very high, the potential of a few exclusions to affect AYP results suggests that states consider using non-matched students' status results rather than excluding them from otherwise growth-only applications.

Reporting option #2: Applying both growth and status criteria to all schools' AYP determinations

The GMPP required states to compile student-level data on both status and growth criteria, and these data could be used to classify each school simultaneously on both criteria for AYP-reporting purposes. That is, each school could be classified as making AYP by both growth and status, by growth-only, by status-only, by a mix of status and growth, or by safe-harbor only. This would have the advantage of uniquely identifying different sets of schools (those making AYP in terms of both growth and status). A possible hierarchy for reporting purposes would be:

- Met AMO by growth and by status (best)
- Met AMO by growth but not status
- Met AMO by status but not growth
- Met AMO by combining status and growth results, so that status results are used for students not on-track to proficiency
- Did not make AMO (worst)

A disadvantage is that some states (e.g., Delaware and Iowa) classified all students who were proficient using the status criterion as also being on-track to proficiency on the growth criterion, regardless of their prior year scores. This would have the effect of increasing the number of "met by growth and status" schools and would prevent comparisons with states that measured growth for proficient and non-proficient students. However, the transition matrix models used by those states could be extended to include categories above proficiency and thus allow for growth-only inferences for proficient students.

Selecting among alternative growth models in the current policy context

This study has also shown that the types of growth models states select for federal accountability purposes are consequential and raise some potentially difficult theoretical questions for policymakers. Projection models are likely to be more accurate than transition matrix or trajectory models in terms of predicting students' future attainment of proficiency targets but the simpler trajectory and transition matrix models may provide clearer guidance to schools, teachers, students, and their parents.

The projection model is explicitly designed to provide probabilistic predictions whereas the other models do not. As a probabilistic estimator, the projection model carries a measure of uncertainty for each student's predicted score. An important issue illustrated by the Ohio model is whether to adjust a student's predicted score for the uncertainty and, if so, to what extent. The adjustment used by Ohio (adding two standard error units to each student's predicted score) was selected in order to make it highly unlikely that a student who actually was on-track was misclassified as not on-track. However, the adjustment also had the effect of classifying a much higher percentage of non-proficient students as on-track than the other pilot states. Careful consideration of the trade-off between false-negatives and false-positives is needed when adjustments for uncertainty are made to statistically derived growth model results such as those from projection models.

Alternatives to proficiency-based growth targets

The growth targets identified by the growth models are tied to the *ESEA* goal of universal proficiency by 2014. This means that substantial individual student performance improvements that do not reach the growth models' proficiency targets (generally within three or four years, or by grades 8 or 9) or subgroup- and school-aggregate student performance improvements that do not reach the 2014-driven proficiency targets (AMOs) are not recognized by the GMPP growth models. However, while not permitted under current *ESEA* regulations, growth models could be adapted to other targets with the result that more students and schools would be identified as making adequate progress than is currently the case. Such other targets generally involve use of a more finely graduated set of performance outcomes than proficiency or on-track to proficiency and moving away from the goal of universal proficiency by 2014.

References

Betebenner, D. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, 28(4), 42–51.

Booher-Jennings , J. (2005). "Below the Bubble: 'Educational Triage' and the Texas Accountability System." *American Educational Research Journal*, 42: 231–268.

Chudowsky, N., Chudowsky, V., and Kober, N. (2009). Is the Emphasis on "Proficiency" Shortchanging Higher- and Lower-Achieving Students? Center for Education Policy, Washington, D.C. Retrieved July 2009 from http://www.cep-dc.org/index.cfm?fuseaction=document_ext.showDocumentByID&nodeID=1&DocumentID=280

Council of Chief State School Officers (CCSSO) (2008). Implementer's Guide to Growth. Washington, D.C. Retrieved May 2008 from http://www.ccsso.org/content/pdfs/IGG%20Final%20AP.pdf.

Ho, A. D., Lewis, D. M., and Farris, J. L. M. (2009). The dependence of growth-model results on proficiency cut scores. *Educational Measurement: Issues and Practice*, 28(4), 15–26.

Neal, D., and Schantenback, D. (August 2007). National Bureau of Economic Research, Working Paper 13239. *Left Behind By Design: Proficiency Counts and Test-Based Accountability.* Cambridge, MA. Retrieved September 2008 from http://www.nber.org/papers/w13293.pdf.

U.S. Department of Education, Office of Elementary and Secondary Education (Jan. 15, 2009). E*valuation of the 2005–06 Growth Model Pilot Program*. Washington, D.C. Retrieved Feb. 2009 from http://www.ed.gov/admins/lead/account/growthmodel/gmeval0109.doc.

U.S. Department of Education, Office of Elementary and Secondary Education (July 2007). *No Child Left Behind* Fact Sheet: "Growth Models: Ensuring Grade-Level Proficiency for All Students by 2014." Washington, D.C. Retrieved May 2008 from http://www.ed.gov/admins/lead/account/growthmodel/proficiency.pdf.

U.S. Department of Education (Jan. 25, 2006). "Peer Review Guidance for the *NCLB* Growth Model Pilot Applications." Washington, D.C. Retrieved May 2008 from http://www.ed.gov/policy/elsec/guid/growthmodelguidance.pdf.

U.S. Department of Education, Office of the Secretary, Office of Public Affairs (2003). *No Child Left Behind: A Parent's Guide*. Washington, D.C. Retrieved May 2008 from http://www.ed.gov/parents/academic/involve/nclbguide/parentsguide.pdf.

U.S. Department of Education (2008). "U.S. Secretary of Education Margaret Spellings Approves Additional Growth Model Pilots for 2007–2008 School Year," press release, June 10, 2008. Retrieved June 2008 from http://www.ed.gov/news/pressreleases/2008/06/06102008.html.

U.S. Department of Education, Office of Communications and Outreach (2009). "U.S. Secretary of Education Margaret Spellings Approves Additional Growth Model Pilots for 2008–2009 School Year", press release, Jan. 8, 2009. Retrieved January 2009 from http://www.ed.gov/news/pressreleases/2009/01/01082009a.html.

U.S. Department of Education, Office of Communications and Outreach (2006). "Secretary Spellings Approves Additional Growth Model Pilots for 2006–2007," press release, Nov. 9, 2006. Retrieved June 2008 from http://www.ed.gov/news/pressreleases/2006/11/11092006a.html.

U.S. Department of Education, Office of Communications and Outreach (2005). "Secretary Spellings Announces Growth Model Pilot, Addresses Chief State School Officers' Annual Policy Forum in Richmond," press release, Nov. 18, 2005. Retrieved May 2008 from http://www.ed.gov/news/pressreleases/2005/11/11182005.html.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service (2007). *State and Local Implementation of the No Child Left Behind Act, Volume III—Accountability Under NCLB: Interim Report*, by K. LeFloch, F. Martinez, J. O'Day, B. Stecher, J. Taylor, and A. Cook. Washington, D.C.: Author.

Appendix A:
Comparison of GMPP Growth Models with State Accountability Systems

Of the nine states approved for the GMPP for the 2007–08 school year, six (Arizona, Arkansas, Florida, North Carolina, Ohio, and Tennessee) already had implemented formal measures of student growth for state accountability purposes before their GMPP models were approved and another (Alaska) implemented a state growth model concurrently with starting the in GMPP. Iowa included growth as an optional measure. In comparison, a 2007 survey by the Council of Chief State School Officer's (CCSSO) Accountability Systems and Reporting State Collaborative on Assessment and Student Standards (ASR-SCASS) found that only six of the 23 responding states had fully operational non-AYP growth models and fewer than half calculated any form of growth at all.[39] This suggests that states with growth modeling experience were better positioned to apply for and receive approval to participate in the federal program. In fact, the two states first approved for GMPP had the most prior experience using student progress to grade schools. Exhibit A.1 provides brief descriptions of each state's pre-GMPP growth model.

[39] See http://edmeasure.com/ASR/doku.php?id=asr:survey_results (retrieved May 2008). A GAO survey in March 2006 found that 26 of 50 states were using growth models, of which seven measured individual student growth (GAO–06–948T). Three of those states are in the GMPP, and our review of documentation suggests that three other GMPP states were using (or about to use) student growth information for state accountability purposes.

Exhibit A.1
Overview of Growth Components of State Accountability Systems

State	Non-AYP Growth Model	Year Implemented	Grades and Subjects Covered	Growth Standards Applied	School Progress Measured	Other Features
Alaska	School Performance Incentive Program	2007–08	3–10 Reading/Writing/ Math	Growth is measured using seven categories of performance.	Points are awarded for student movement among performance categories and are averaged to create a school index score.	Four performance levels are used to reward schools according to their growth index scores.
Arizona	Measure of Academic Progress (MAP)	1999–2000, revised 2005–06	4–8 Reading/Math	Growth is measured using annual differences in scale scores and is averaged to calculate school growth.	Points are awarded for meeting expected growth (by quartile) and are added to other measures of school performance (AZ LEARNS).	Expected growth is calculated using a value-added model adjusted to ensure proficiency by seventh grade and controlling for mobility or ceiling effects.
Arkansas	Act 35 Annual School Ratings System	2003–04, growth added in 2008–09	3–11 Reading/Writing/ Math	Student growth is based on changes in eight performance levels across adjacent years.	Value points are added if a student moves up a performance category and are lost if a student moves down a category.	Schools are evaluated using five performance levels for both "improvement" (Category One) and "status" (Category Two).
Delaware	n/a	n/a	n/a	n/a	n/a	n/a
Florida	Grading Florida Public Schools	2001–02	3–10 Reading/Math/ Science/Writing	Growth is measured by movement among five performance levels.	Grades schools by awarding points for the number of students meeting proficiency standards plus those making annual learning gains.	Points also are awarded for students who maintain high performance, for the percent of students in the lowest quartile making gains, and if half or more of such students make gains.
Iowa	*Optional*	n/a	K–12 Reading/Math/ Social Studies/ Science	Growth for student profiles is measured using changes in ITBS and ITED scores.	n/a	n/a

Exhibit A.1 continues next page

Exhibit A.1
Overview of Growth Components of State Accountability Systems
continued from previous page

State	Non-AYP Growth Model	Year Implemented	Grades and Subjects Covered	Growth Standards Applied	School Progress Measured	Other Features
North Carolina	ABCs Growth Model	1996–97, revised 2005–06	K–8, HS Reading/Math/ Science	Growth is the average of standardized scores from two previous years adjusted for the tendency of students with extremely high or low initial score to score closer to the mean on the next assessment ("regression to the mean") and is expected to be "0" or above, indicating at least one year's worth of gain.	A mean growth score is calculated for the school, though schools can also meet expected growth if 60 percent or more of students test at grade-level proficiency.	Schools get credit for "high growth" if students who are stable or growing outnumber those declining by at least three to two [or a ratio of 3:2].
Ohio	Value-Added Assessment	2006–07	4–8, plans for K–10 Reading/Math	Growth is measured as mean normal curve equivalent (NCE) gains and is categorized as meeting or as being a standard error "above" or "below" the "expected" gain.	LRC designations can be raised if schools are above expected growth for at least two years and lowered if below expectations for at least three years.	Value-added is based on the 2006–07 distribution of test scores and will be used for accountability beginning in 2007–08.
Tennessee	Tennessee Value-Added Assessment System (TVAAS)	1992–93	3–8 Reading/Math/ Language/ Science/Social Science	Growth in TCAP results is measured as mean normal curve equivalent (NCE) gains relative to a state growth standard (1998 Terra Nova) or a state three-year average.	School reports include average student gains by grade and whether achievement level meet growth standards or are within one, two, or more standard deviations.	Specific school effects also are estimated using up to five years of student data to predict scores for the average student in a given school; teacher effect data is restricted access.

However, these state accountability growth models had to be modified in significant ways for use in GMPP. One basic difference between models approved under *ESEA* and those used for state accountability purposes is the definition of expected growth. The seven core principles require that the *ESEA* growth models place limits on the amount of time students have to reach proficiency, usually by the last grade tested as applied, resulting in growth expectations that put a student "on-track" to meet grade-level standards in a few years. Most state growth models do not require students to be on-track to proficiency but, for example, simply expect students to exhibit a year's worth of growth per year of instruction. The principle of universal proficiency by 2014 also translates into rising AMOs, a requirement that is absent from state accountability systems. The goal of states in most cases is to determine whether individual schools and the system as a whole are improving rather than to see if a given student will be proficient in a few years or if a school will have all students proficient by 2014.

Another major difference is that growth components of state accountability systems typically calculate an average for the entire school instead of basing school performance on the results of each subgroup separately and may average over subjects as well. This means that a school can still make adequate growth for state accountability purposes even if one or more subgroups fails to grow or if the school exhibits strong growth in one subject but not another. Of course, this does not mean that growth results for subgroups and subjects are not reported, and states often measure growth for more than reading and math. The goal is to create a composite growth score that can then be combined with other measures of school performance (including AYP) both for reporting purposes, such as a "grade" on school report cards, and for rewarding schools that meet or exceed state standards.

The result of these differences is that more schools tend to make adequate growth for state accountability purposes than would make AYP under the GMPP. For example, 779 of the 1,298 (or 60 percent) of the North Carolina schools that failed to make AYP in 2006–07 made expected growth under the state guidelines. Of these, 187 (or 14 percent) met the state standards for high academic growth. The federal growth model used by North Carolina sets the bar higher for both expected growth and percent proficient (AMO) than does the growth model used for state accountability, suggesting that the universal proficiency requirement is driving most if not all of this difference. Florida also drops the proficiency deadline in its grading system, although the transition matrix model used sets expectations that may differ from a single year's worth of growth.[40] The school itself can make adequate growth if half or more of low-achievers meet expected growth, which is a different standard than the AMO used for AYP purposes. Thus about 84 percent of Florida schools made adequate growth in both reading and math for 2006–07, compared to the 34 percent of schools that made AYP under the GMPP.

Offsetting this tendency of the 2014 deadline to lower the number of schools making adequate growth is that state accountability systems do not give schools credit if enough students grow to push a subgroup or school over the proficiency threshold as is done in the status-plus-growth

[40] Note that students who score in the two lowest FCAT achievement levels still can make expected growth if they make a year's worth of growth as determined by FCAT developmental scores.

models implemented under the GMPP.[41] This means that schools classified as making AYP by growth may not be showing much growth for the student body as a whole. Thus of the 12 North Carolina schools that made AYP by growth in 2006–07, only three met the state's high growth standard and one did not make the lower standard of expected growth. Similarly, 10 of the 19 Tennessee schools that made AYP by growth in 2006–07 grew at a rate below the state three-year average for math, while 14 schools grew at a below average rate for reading. One school even experienced negative growth for both reading and math. Growth models are only one component of federal accountability under the GMPP and thus do not provide much information about whether a school is growing as currently reported. However, results from growth measures used for state accountability purposes suggest that many more schools would make AYP if the first of the seven core principles of the *ESEA* project was relaxed.

[41] Note that states often give growth credit to schools for *maintaining* proficiency, which is similar to a growth-plus-status model, and may grade schools on proficiency status criteria as well.

Appendix B:
State GMPP Model Summaries

The core of each growth model is the method of determining whether a student is "on-track" to being proficient. While states' methods can be categorized as one of three types of growth models (trajectory, transition matrix, or projection), each state's method has unique features. This appendix offers additional details about the growth models used in Alaska, Arizona, Arkansas, Delaware, Florida, North Carolina, and Tennessee. All the relevant detail of Iowa's growth model is included in the main text. Ohio used the same growth model as Tennessee and the latter's summary is sufficient for both states with the important difference that Ohio adjusted predicted scores by the standard errors of the predictions (see Ohio summary above).

Alaska

Alaska uses a trajectory model that calculates growth for students in grades 4 through 6, grade 8 and grade 9, using results of the Alaska Standards Based Assessment Test mathematics and language arts.

Proficiency levels for these tests are scaled so that thresholds are 300 for math and 600 for literacy within each grade. The literacy score is cutoff at 600 because it is a combination of reading and writing, each of which has a cutoff of 300 points. For this summary we will consider a generic cutoff score of 300 for each grade.

Students are always classified by the traditional status model in grades 3 and 7. In grades 4 through 6 it is possible to use the growth model. In the Alaska growth model the first year a student is below proficient is considered his or her base year, and growth targets for subsequent years are calculated by evenly dividing the difference between the base year score and the cutoff in grade 7. In grades 8 and 9, the first year a student is below proficient is considered the base year, and growth targets for subsequent years are calculated by evenly dividing the difference between the base year score and the cutoff in grade 10.

For example, if a student scores below proficient in some grade q (between grades 3 and 6), growth targets for each subsequent grade are a function of his or her base score in grade q, Y_q, plus the difference between this score and 300 weighted fraction of the years that have elapsed between the previous grade and grade 7. Thus, a student in grades 4 through 6 is considered to be on-track toward proficiency in grade k if his or her score, Y_k, is greater than or equal to this function,

$$Y_k \geq Y_q + \left(300 - Y_q\right) \times \left(\left(k - q\right) / \left(7 - q\right)\right)$$, assuming $k > q$ and $Y_q < 300$.

For students who score below proficient in some grade q between grades 7 and 10, growth targets for each subsequent grade are a similar function of the student's base score in grade q,

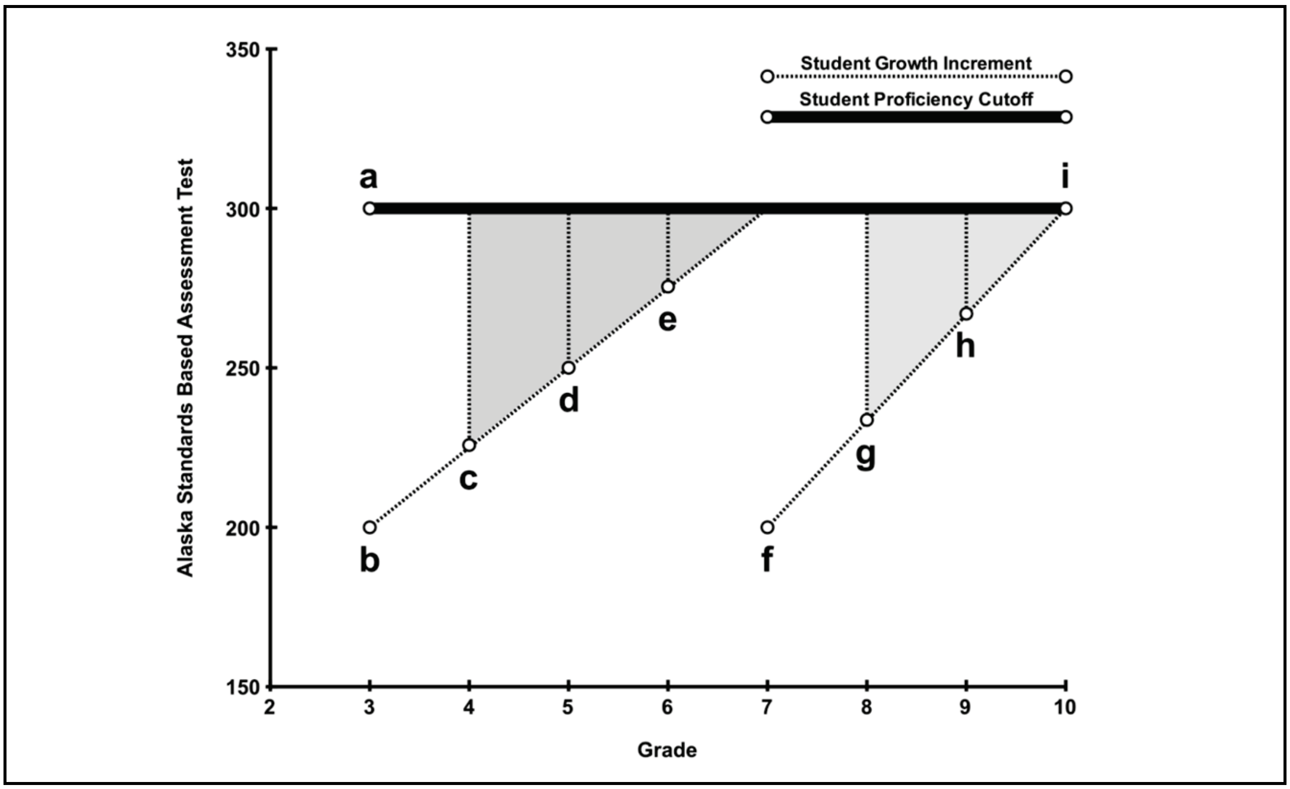
Y_q, plus the difference between this score and 300 weighted fraction of the years that have elapsed between the previous grade and grade 10. Thus, a student in grades 8 or 9 is considered to be on-track toward proficiency in grade k if his or her score, Y_k, is greater than or equal to this function, or,

$$Y_k \geq Y_q + \left(300 - Y_q\right) \times \left(\left(k - q\right) / \left(10 - q\right)\right)$$, assuming k > q and $Y_q < 300$.

These targets, once established by the baseline year, remain the growth targets for subsequent years.

The illustration in exhibit B.1 below showcases this process. This figure shows the proficiency standard for each grade (300) along the solid line connecting points a to i. The shaded polygons illustrate the necessary growth between years to be on-track toward proficiency for a set of hypothetical students.

Exhibit B.1
Illustration of Alaska's Method for Determining Whether a Student Is On-Track Toward Proficiency



Label	a	b	c	d	e	f	g	h	i
Grade	3	3	4	5	6	7	8	9	10
Score	300	200	226	250	275	200	234	267	300

Student Z is below proficient in the third grade, he scored 200 (point b) instead of the proficiency standard of 300. In applying the formulas above, Alaska's growth model would set a new target score of 226 in the fourth grade (point c), 250 in grade 5 (point d), and 275 in grade 6 (point e), to be considered on-track for those grades. While that student cannot be used for growth model AMO calculations in the third grade, that student will be expected to make those targets calculated for subsequent years to be considered as on-track in grades 4 through 6. Similarly, student X is below proficient in the seventh grade, he scored 200 (point f) instead of the proficiency standard of 300. In applying the formulas above, Alaska's growth model would set new target scores of 234 in the eighth grade (point g), and 264 in grade 9 (point h), to be considered on-track for those grades. While that student cannot be used for growth model AMO calculations in the seventh grade, that student will be expected to make those targets calculated for subsequent years to be considered as on-track in grades 8 and 9.

Arizona

Arizona uses a trajectory model to set growth targets for students in grades 3 through 8 using Arizona's Instrument to Measure Standards (AIMS) test. The score cutoffs for reading and math are presented in Exhibit B.2.

Arizona, like other states using trajectory models, sets growth targets for students who are not proficient based on a baseline score. Growth targets are set by determining the necessary improvement to reach proficiency and setting evenly spaced improvement targets over a given set of years. In Arizona, growth targets are set based on the first non-proficient score. Growth targets calculated from a baseline score in K through third-grade require the student to reach proficiency by the sixth grade. Growth targets calculated based on a fourth grade score require the student to be proficient by seventh grade. Growth targets calculated based on scores in grades 5, 6, and 7 require the student to be proficient by the eighth grade. Growth targets are reset each year based on student performance and movement in and out of the Arizona public school system. However, the specific grade required to reach proficiency does not change when targets are reset.

Exhibit B.2
Arizona Proficiency Standards Test Cutoffs for Grades 3 Through 8

Grade	Reading Cutoff	Math Cutoff
3	431	420
4	450	448
5	468	476
6	478	496
7	489	517
8	499	537

While each student is assigned growth targets based on actual performance, Arizona is unique compared to other growth model pilot states in that schools are evaluated on each student's estimated score instead of students' actual scores. However, in calculating the proportion of students within a school who are considered to be on-track toward proficiency, students are classified based on the lower bound of their predicted scores. These predicted scores are generated from a statewide statistical model which fit current year scores to scores from the previous years.

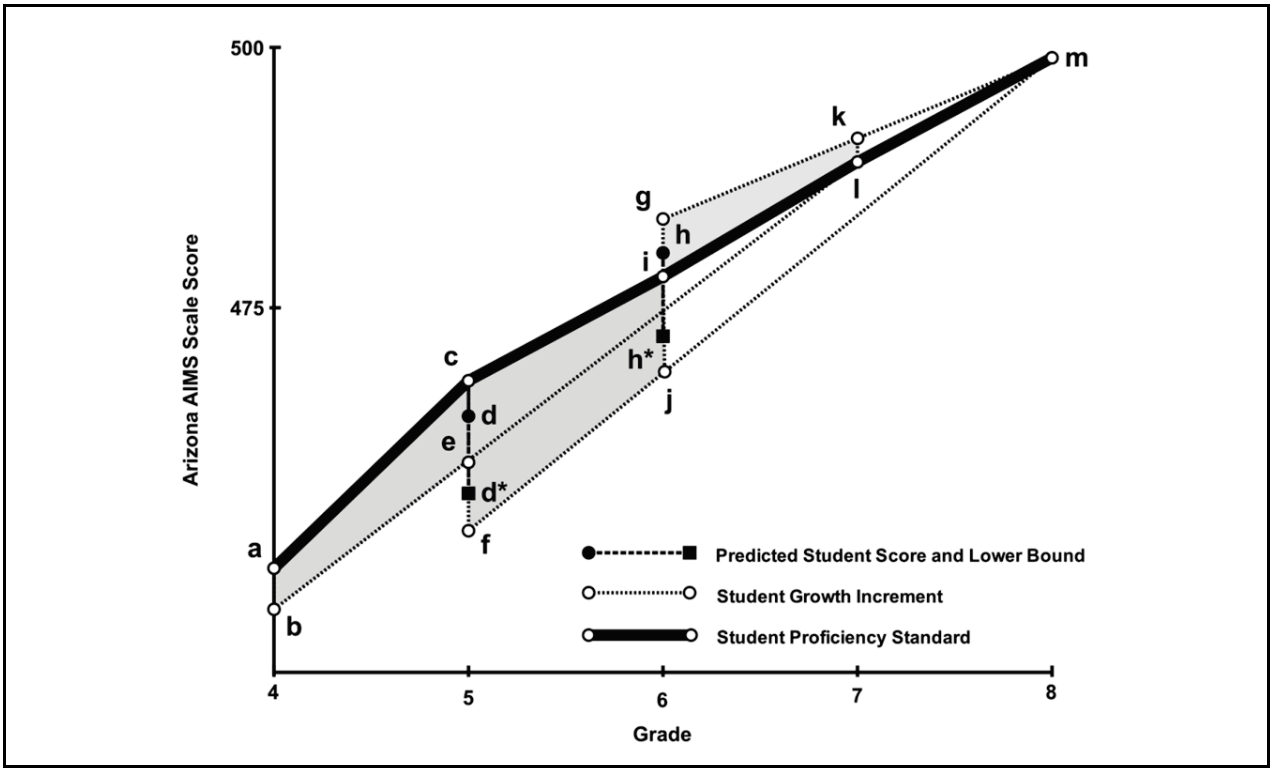
Arizona estimates a regression model predicting the current year scores of all students by their previous year's score. Let Y_{it} be the score from student i in grade k in school j at the current year t, $Y_{it} = \alpha + \beta Y_{it-1} + e_i$. From this model, Arizona predicts each student's current score, $Y^*_{it} = \alpha + \beta Y_{it-1}$. There is a statewide standard error, $SE_{Y^*_t}$, associated with each year's predicted score, which typically equals about 4 or 5 scale score points. The score used to evaluates a student's growth, Y^*_{it}, is then a lower bound of a confidence interval estimated around the predicted score, or $Y^{Lower}_{it} = Y^*_{it} - \left(1.96 \times SE_{Y^*_t}\right)$. As a result, each student is evaluated not on his or her actual score, but on this lower bound of an estimated predicted score.

The illustration in exhibit B.3 below showcases this process. This figure shows the proficiency standards for grades 4, 5, 6, 7, and 8 along the solid line connecting points a, c, i, l, and m. The shaded polygons illustrate the necessary growth between years to be on-track toward proficiency for a of hypothetical student.

Student Z is below proficient in the fourth grade, he scored 446 (point b) instead of the proficiency standard of 450 (point a). Arizona's growth model would set a new target score of 460 in the fifth grade (point e) to be considered on-track. Student Z then scored exactly 454 in the fifth grade (point f), which is below the necessary growth target. However, the student predicted score is 465 (point d), which is above the growth target. Yet, the school will be unable to count that student as on-track because the lower bound of the predicted confidence interval is 457 (point d*), still better than the actual score, but also still below the growth target. The student actual fifth grade score is then used to estimate a new growth target of 469 in the sixth grade (point j).

Student Z then did well in the sixth grade, scoring 484 (point g), well above the normal cutoff of 478 (point i). However, that student's predicted score is 480 (point h) with a lower bound 472 (point h*). Because the lower bound is still greater than the growth target, that student can then be counted as on-track. Under a growth only model, that student's growth target for seventh grade would be higher than the normal cutoff (point k instead of point l) because that student's actual sixth-grade score was greater than the normal cutoff as well.

Exhibit B.3
Illustration of Arizona's Method for Determining Whether a Student Is On-Track Toward Proficiency



Label	a	b	c	e	d	d*	f	g	h	h*	i	j	k	l	m
Grade	4	4	5	5	5	5	5	6	6	6	6	6	7	7	8
Score	450	446	468	460	465	457	454	484	480	472	478	469	491	489	499

Arkansas

Arkansas uses a trajectory model that calculates growth for students in grades 4 through 7 using results of the Arkansas Benchmark Exams for mathematics and literacy, which are administered in grades 3 through 8 (plus grade 11 for literacy).

Proficiency levels for these vertically scaled exams are set for each grade, and growth targets are based on the annual exam score increment needed to reach the proficiency standard in eighth grade. These proficiency standards are presented in table B.4 below.

Exhibit B.4
Arkansas Benchmark Exam Proficiency Standards for Grades 3 Through 8

Grade	Proficiency Standard
3	500
4	559
5	604
6	641
7	673
8	700

If a student does not achieve proficiency in a base year, the growth targets necessary to be considered on-track toward proficiency are calculated based on the total increment between a student's base score and the proficiency standard in eighth grade (700). However, instead of setting growth targets by evenly dividing the total increment between the base score and 700, the increments are proportioned so that the growth increments follow a curve that matches the concave nature of the typical proficiency standards.

This means that, for example, instead of setting the growth increment from third to fourth grade to be a quarter (0.25) of the total necessary improvement to reach 700 by eighth grade, the growth increment for fourth grade is set to slightly more than a quarter (0.295) of the total necessary improvement to reach 700 by eighth grade. Thus, the annual increment that a student must attain in order to be classified as on-track to proficiency is calculated using grade-specific growth target multipliers. The multiplier for any grade "k" is simply a ratio of the difference between the current and previous year's proficiency standards, P_k and P_{k-1}, to the difference between the proficiency standard for eighth grade and the previous year, or,

$$\left(P_k - P_{k-1}\right) / \left(700 - P_{k-1}\right).$$

These multipliers are presented in Exhibit B.5.

Exhibit B.5
Arkansas Growth Target Multipliers for Grades 3 Through 7

Grade	Multiplier
4	0.295
5	0.319
6	0.385
7	0.542

Another method of illustrating a specific growth increment is to use the following formula that calculates the minimum score necessary to be considered on-track toward proficiency. For any student in grade k, the minimum score necessary to be considered "on-track" is his or her previous score, Y_{k-1}, plus the difference between 700 and his or her previous score, $700-Y_{k-1}$, times the multipliers outlined above. Thus, if the current score, Y_k, is greater than or equal to this function, or,

$$Y_k \geq Y_{k-1} + \left(\left(\left(P_k - P_{k-1} \right) / \left(700 - P_{k-1} \right) \right) \times \left(700 - Y_{k-1} \right) \right),$$

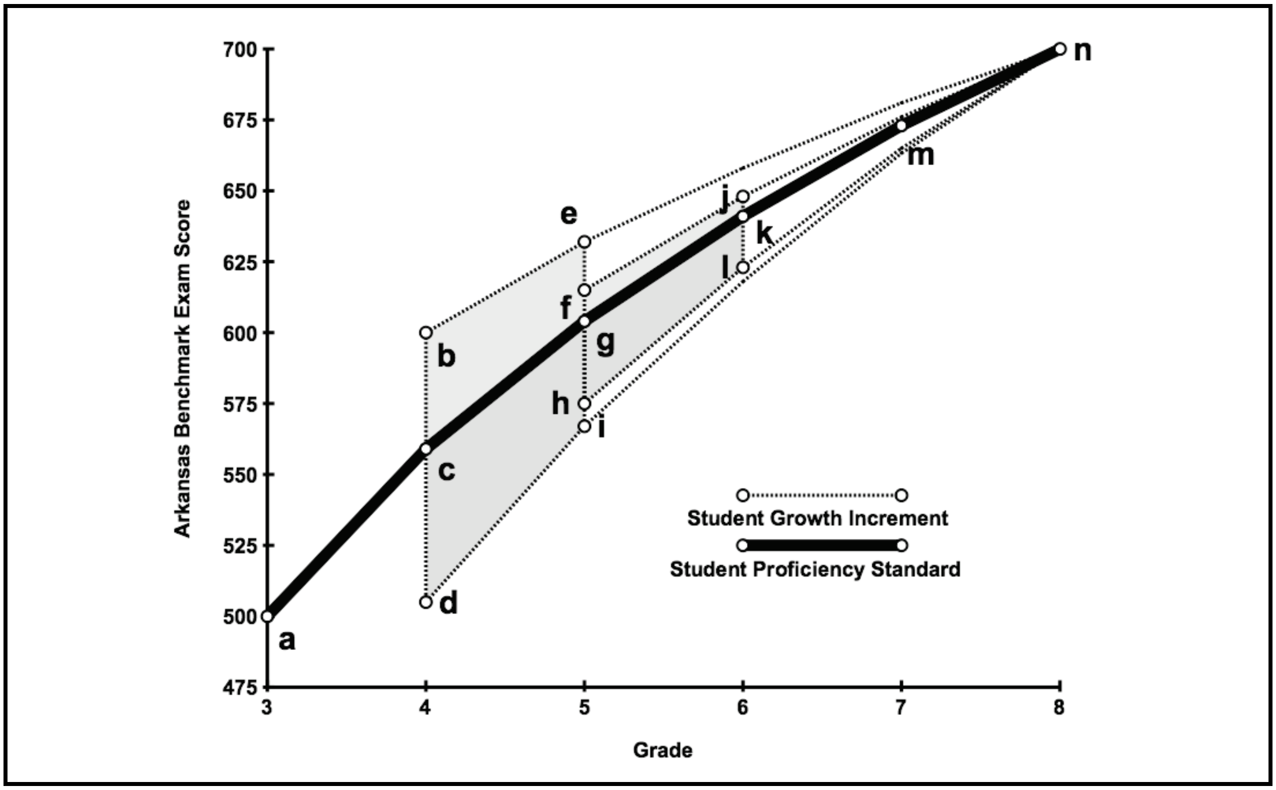
that student is considered to be on-track toward proficiency. From this formula, we see that this model resets the growth target every year rather than setting a series of annual targets based on the first below-proficient exam score. This formula can also determine minimum scores even for students who score above the proficiency threshold. Thus, students can be identified who are not improving toward the eighth-grade standard of 700 even though they are currently proficient.

The illustration in exhibit B.6 below showcases each of these features. This figure shows the proficiency standards for each grade along the solid line connecting points a, c, g, k, m and n. The shaded polygons illustrate the necessary growth between years to be on-track toward proficiency for a set of hypothetical students.

Student Z is below proficient in the fourth grade, he scored 505 (point d) instead of the proficiency standard of 559 (point c). In applying the formulas above, Arkansas' growth model would set a new target score of 567 in the fifth grade (point i) to be considered on-track. If student Z scored exactly 567 in the fifth grade, then his growth targets would follow the dotted line from point i to point n. In addition, imagine that student Z in fact did better than scoring 567 in the fifth grade—let us say that he scored 575 (point h). Then, that student's growth target for sixth grade would be reset and instead would be 623 (point l). Note that because student Z did better than his growth target in the fifth grade, his new growth target for the sixth grade is slightly higher—point l is above the dotted line from point i to point n.

Another student, student X, scored well above the standard in fourth grade at 600 (point b). By applying the same formula, the Arkansas growth model would set a fifth-grade target score of 632 (point e), which is likewise well above the proficiency standard of 604 (point g). Next, imagine that student X scored lower than his target score in fifth grade—let us say he scored 615 (point f). Student X scored above proficient in fifth grade, but because is score is lower than his target score generated from fourth grade, he is not considered to be on-track toward proficiency. The Arkansas model does reset scores each year, and so we can see that his target for sixth grade is set at 648 (point j), which is closer to the proficiency standard of 641 (point k). Note that his sixth-grade target is below the dotted line from point e to point n. This shows that his new target score is reflecting that he did not score as high in the fifth grade as he did in the fourth grade. Still, he would need to score at least 648 in the sixth grade (point j), which is more than the normal sixth-grade proficiency standard of 641 (point k).

Exhibit B.6
Illustration of Arkansas' Method for Determining Whether a Student Is On-Track Toward Proficiency



Label	a	b	c	d	e	f	g	h	i	j	k	l	m	n
Grade	3	4	4	4	5	5	5	5	5	6	6	6	7	8
Score	500	600	559	505	632	615	604	575	567	648	641	623	673	700

Delaware

Delaware's transition matrix model uses a "value table" method for AYP determinations that assigns points for students depending on the type and extent of changes between the performance levels (see Exhibit 6 in the report). Points in the value table increase with both the magnitude of change in level of proficiency, and the overall level of proficiency. As a result, a student scoring toward the bottom during his previous grade receives fewer points for moving up one level than he would if he moved up several levels. All students who surpass their grade level cutoff receive the maximum points regardless of how high they scored or whether their underlying scores actually declined from the prior year. From these points assigned to students, schools meet their AMO if the average number of points per student meets or exceeds a target number of points set by Delaware for particular subjects and grades.

Delaware's point system is designed to be an analogue to the AMO system used for status models by giving partial weight to students who are below proficient but still making progress.

In fact, the target number of points that Delaware sets is simply the standard AMO that is rescaled to the point system.

For example, in the 2006–07 school year the status model AMO for reading was 68 percent proficient. For that same year, the target number of points that a school's students needed to gain on average was 204. If we divide the target points by the maximum number of points a student can receive, we return to the AMO (204/300 = 0.68). To illustrate how a school meeting the AMO under the status model would perform under the growth model, imagine two schools with 100 students, school Z and school X (see exhibit B.7 below).

In the first school, school Z, exactly 68 out of 100 students (68 percent) were proficient in reading. That school meets the AMO of 68 percent exactly. Let us assume that in this school all students who were below proficient did not move up any categories from their previous grade. Then, under Delaware's growth model, that school would receive

$$68 \times 300 = 20,400 \text{ points,}$$

out of a total possible

$$100 \times 300 = 30,000 \text{ points.}$$

The average number of points per student would be

$$\frac{20,400}{100} = 204,$$

which is exactly the target number of points a school needs to make their AMO under the growth model, and the number of achieved points, 20,400, is exactly 68 percent (the AMO) of the total possible 30,000 points. Thus school Z would make their AMO under both the status model and the growth model.

To illustrate how a school that would not meet its AMO under the status model, but would make its AMO under the growth model, consider another school X which also has 100 students, but only 30 students are proficient. Under the status model, only 30 percent of the students are proficient, which is well below the target AMO of 68 percent. Yet, the school does have many students who are improving. Under the growth model, the proficient students account for only

$$30 \times 300 = 9,000 \text{ points,}$$

which is only 30 percent of the total possible 30,000 points. Remember, though, school X has students who are improving, including 30 students who moved from PL 1A to PL 1B for

$$30 \times 150 = 4,500 \text{ points,}$$

and another 15 students moved from PL 1B to PL 2A for

$$15 \times 175 = 2{,}625 \text{ points,}$$

and another 15 students moved from PL 1B to PL 2B for

$$15 \times 225 = 3{,}375 \text{ points,}$$

and 5 students who did not improve, who stayed at PL1A for

$$5 \times 0 = 0 \text{ points.}$$

That school would total 20,500 points out of a possible 30,000 points. That would be 68.3 percent of the total possible points, greater than the AMO for reading that year. The average number of points per student would be 205, which is larger than the 204 points necessary to meet the AMO. Under the growth model, school X would meet the AMO.

This means that Delaware's status model is actually like a simple transition matrix in which all students who are proficient are given the maximum number of points, and all student who are below proficient are given no points at all. The growth model, then, gives students who improve a partial number of points. The table in Exhibit B.7 can be used to estimate the proportion of the total points each improvement is worth. Another way to think about Delaware's growth model is that a student who moves from the first level, PL 1A, to the second level, PL 1B, is given half (0.50) the weight of a student who is proficient when the percent proficient is calculated. This means that in Delaware's transition matrix system, two students who move from the first to the second level equal one student who is proficient.

Exhibit B.7
Example Comparison of Two Hypothetical Schools in Delaware to Illustrate Growth Model

Year 1	Year 2	Points Per Student	Proportion of Total Possible Points	School Z		School X	
				Number of Students Per Cell	Points Per Cell	Number of Students Per Cell	Points Per Cell
PL 1A	PL 1A	0	0.00	0	0	5	0
PL 1A	PL 1B	150	0.50	0	0	30	4,500
PL 1A	PL 2A	225	0.75	0	0	0	0
PL 1A	PL 2B	250	0.83	0	0	0	0
PL 1A	Proficient	300	1.00	0	0	0	0
PL 1B	PL 1A	0	0.00	0	0	0	0
PL 1B	PL 1B	0	0.00	0	0	0	0
PL 1B	PL 2A	175	0.58	0	0	15	2,625
PL 1B	PL 2B	225	0.75	0	0	15	3,375
PL 1B	Proficient	300	1.00	0	0	0	0
PL 2A	PL 1A	0	0.00	0	0	0	0
PL 2A	PL 1B	0	0.00	0	0	0	0
PL 2A	PL 2A	0	0.00	0	0	0	0
PL 2A	PL 2B	200	0.67	0	0	5	1,000
PL 2A	Proficient	300	1.00	0	0	0	0
PL 2B	PL 1A	0	0.00	0	0	0	0
PL 2B	PL 1B	0	0.00	0	0	0	0
PL 2B	PL 2A	0	0.00	0	0	0	0
PL 2B	PL 2B	0	0.00	32	0	0	0
PL 2B	Proficient	300	1.00	0	0	0	0
Proficient	PL 1A	0	0.00	0	0	0	0
Proficient	PL 1B	0	0.00	0	0	0	0
Proficient	PL 2A	0	0.00	0	0	0	0
Proficient	PL 2B	0	0.00	0	0	0	0
Proficient	Proficient	300	1.00	68	20,400	30	9,000
Total				100	20,400	100	20,500
Percent Proficient Year 2				68%		30%	
Average Points					204		205
Meet AMO by Status (68% proficient)?				Yes		No	
Meet AMO by Growth (204 Average Points)?				Yes		Yes	

Florida

Florida's growth model applies to students in grades 3 through 10 using the Developmental Scale Scores (DSS) from the Florida Comprehensive Assessment Tests (FCAT) for mathematics and reading. The state uses a trajectory model that bases growth targets on the score required for proficiency three years after the first year tested (usually, grade 3). To be counted as proficient, a student who was not proficient at the baseline must close the gap between his baseline score and the proficiency cutoff three grades later by one-third each year. This means that a student can only make proficiency by growth for two years because the gap must be fully closed in the third year, (i.e., a student must meet or exceed the minimum proficiency score in year 3). A student first enrolled in grade 9 must close the gap by half to be counted as "on-track." The cutoffs for grades 3 through 7 are presented in Exhibit B.8.

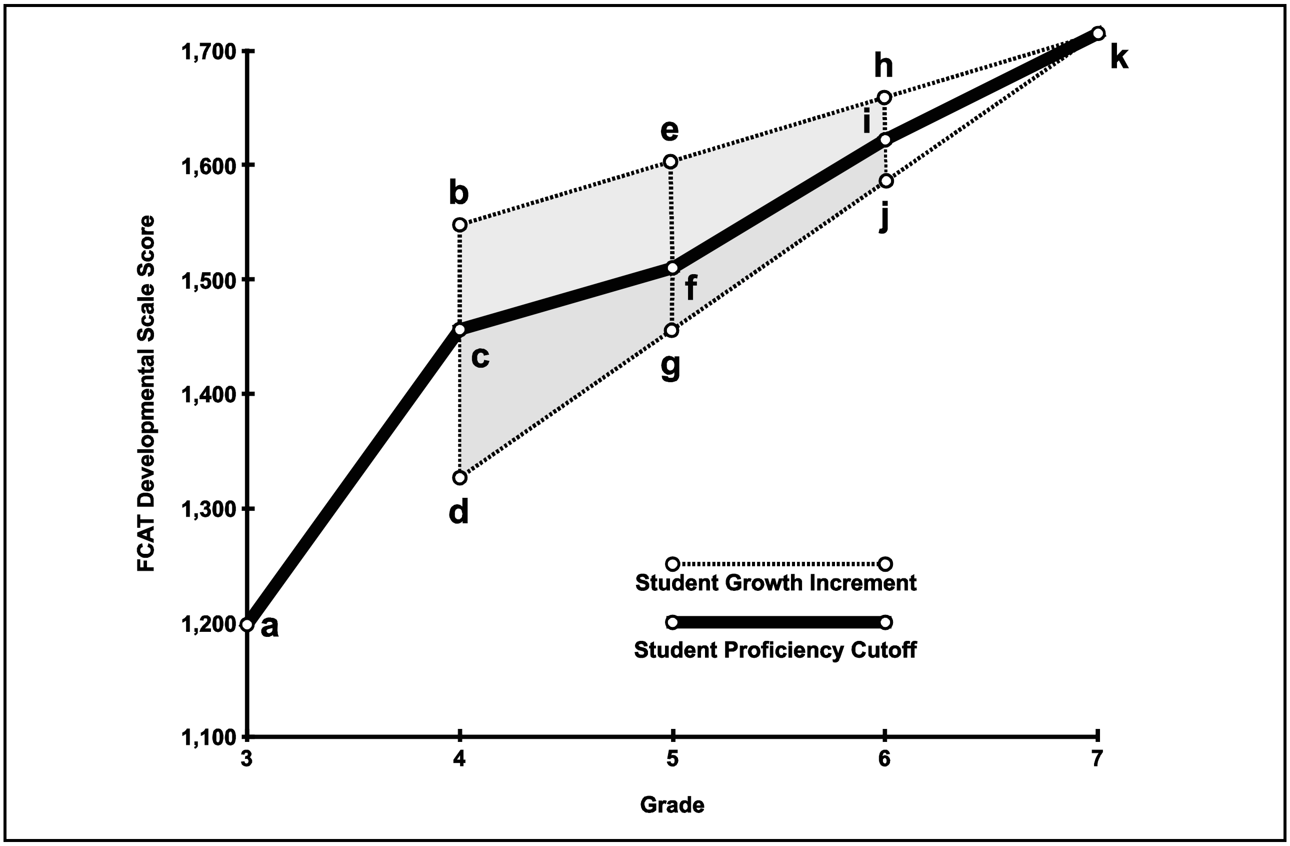
Exhibit B.8
Florida's Cutoff Developmental Scale Scores to Be Considered Proficient on the FCAT for Grades 3 Through 7

Grade	DSS Cutoffs for Reading	DSS Cutoffs for Math
3	1,198	1,269
4	1,456	1,444
5	1,510	1,632
6	1,622	1,692
7	1,715	1,786

The illustration in exhibit B.9 below showcases this process for reading. The bold line connecting points a, c, f, i, and k represent the FCAT DSS score cutoffs for reading for grades 3 through 7. The shaded polygons represent the growth increments necessary to be considered on-track toward proficiency. To illustrate the properties of Florida's growth model, let us revisit student Z and student X.

Student Z is below proficient. He scored 1,326 in fourth grade (point d), which is below the grade-specific cutoff of 1,456. Florida's growth model would then set growth targets for the following two years, fifth grade and sixth grade, by dividing the difference between his fourth-grade score and the seventh-grade cutoff into thirds. In other words, he would be expected to close the gap between his fourth-grade score and the seventh-grade cut-off by a third during the first year after he was not proficient (fifth grade). In turn, he would be expected to close the gap by two-thirds after two years of not being proficient (sixth grade). He would then be expected to be proficient after three years (seventh grade). Geometrically, this is drawing a straight line from point d at grade 4 to point k at grade 7. This line then sets growth targets at point g and point j for grades 5 and 6, respectively. Thus, the growth target for student Z in fifth grade is 1,455 (point g) and 1,586 in sixth grade (point j).

Exhibit B.9
Illustration of Florida's Method for Determining Whether a Student Is On-Track Toward Proficiency in Reading



Label	a	b	c	d	e	f	g	h	i	j	k
Grade	3	4	4	4	5	5	5	6	6	6	7
Score	1,198	1,548	1,456	1,326	1,603	1,510	1,455	1,659	1,622	1,586	1,715

Unlike Arkansas' trajectory model, Florida's method is not intentionally designed as a method for tracking the progress of proficient students. Florida does not scale growth targets to match the pattern of cutoff scores. As a result, this method is very sensitive to the pattern of cutoffs set for each grade. This means that if this method were used to draw growth targets for proficient students, then the magnitude of the higher expectations would fluctuate between grades, depending on grade-by-grade standards. Furthermore, the problem of how far above the cutoff a growth target would be for any one grade would greatly depend on *which* grade was used for the base year.

For example, imagine that student X scored 1,548 (point b) in the fourth grade, which is 92 points above the cutoff of 1,456 (point c). Let us further assume that for some reason, that score was used at a baseline for future growth targets. As a result, that student's growth targets would be represented by the line from point b to point k. By following that line we see that student X would have to score 1,603 points to be considered on-track (point e), which is 93 more points than the cutoff of 1,510 in the fifth grade (point f). He would not be expected to score as many more points in sixth grade, however, because the cutoff scores increase by a large amount between the fifth and sixth grades. In the sixth grade, student X would have to score 1,659 (point h), or only 37 more points than the cutoff of 1,622 (point i). This illustrates that students who score above proficient would be subject to inconsistent expectations because the pattern of cutoff scores in Florida is not linear, and the model is not designed to mimic the pattern of standards from grade to grade or reset expectations.

North Carolina

North Carolina uses a trajectory model to identify growth targets and both End-of-Grade (EOG) tests and End-of-Course (EOC) tests for assessing student progress. The state uses a Standardized Scale Approach (SSA) to growth which uses the normative distribution of student performance in the standard setting year of any test edition as a common basis to build a scale. State documents note that this approach is useful for measuring the growth in student performance from one year to the next and also adapts well to the changes in curriculum and subsequent changes in test editions.

The SSA system uses a time-locked modified z-scale termed a “change scale” or “c-scale.” Thus, the c-scale cut score for proficiency on any given test edition at an individual grade level remains constant for the life of the scale and test edition regardless of the changes in the distribution of test scores that might occur as schools change their instructional methods. The state means and standard deviations from the standard setting year are used indefinitely for any given test.

The 2005–06 school year was the standard setting year for the Mathematics EOG tests at grades 3–8 and the 2002–03 school year was the standard setting year for the Reading EOG tests in grades 3–8. North Carolina performs an equating study to set the achievement level cut scores at the same time the c-scale is built.

The trajectory is built based on the student’s performance either the previous year, or on the third-grade pretest, whichever is appropriate to the grade in which the student first enters the state. Therefore, the following table illustrates the basis for prediction, the targeted test for proficiency, the years of trajectory, and the percent of difference between baseline performance and proficiency expected by the trajectory based on the year the student is first enrolled in the state in a tested grade.

Exhibit B.10
Grades and Tests Used for Trajectory Growth in North Carolina and the Percent of Difference Expected to Be Closed Per Year

Grade of First Enrollment	Test Used as the Basis for Prediction	Test Used as Target for Proficiency	Years to Proficiency	Percent of Difference Closed Per Step	Steps to Proficiency
3	3rd-grade pretest	6th-grade EOG	4	25%	4
4	4th-grade EOG	7th-grade EOG	4	33%	3
5	5th-grade EOG	8th-grade EOG	4	33%	3
6	6th-grade EOG	Algebra I or English I EOC	4	33%	3
7	7th-grade EOG	Algebra I or English I EOC	4	50%	2
8	8th-grade EOG	Algebra I or English I EOC	3	100%	1

The trajectories are built individually by student and separately for reading or mathematics. Therefore, a student will have a trajectory based on their baseline mathematics score and the proficiency cut score for mathematics separate from reading. In the upper grades, Algebra I is the AYP assessment for tenth-grade students and is the trajectory target for math while English I is the trajectory target for reading or language arts.

The following table displays the performance expected of students to be counted as on trajectory for inclusion in the proposed method of comparing school performance to AMO targets.

For a student who enters in third grade and has a grade 3 pretest:

Year in State-tested Grade	Decrease From Baseline Assessment in Performance Discrepancy
1	25% of Original Gap
2	50% of Original Gap
3	75% of Original Gap
4 or more	Student Must Be Proficient

For a student who enters in fourth, fifth, or sixth grade:

Year in State-tested Grade	Decrease From Baseline Assessment in Performance Discrepancy
1	Baseline, Not On Trajectory
2	33% of Original Gap
3	66% of Original Gap
4 or more	Student Must Be Proficient

Therefore, if a subgroup has met its 95 percent participation target but has not met its proficiency target, and the subgroup has met its other academic indicator, the process of incorporating the growth measure would be:

1) First identify if the student has been in membership the full academic year and is both tested and not proficient.
2) These three conditions being met, the number of years the student has been in the state will be determined using the historic files from the state's accountability system.
3) If the student has been in the state (in a tested grade) for four years or more, the student will remain non-proficient for comparison to the annual measurable objectives (AMO). If the student has been in the state public schools three years or less, the correct baseline score will be located (using the table above).
4) The student's performance on the baseline assessment in the subject of interest will be converted to the c-scale.
5) Based on the student's baseline score and proficiency in the target year, a difference will be calculated.
6) The decrease in the difference will be compared against Table 4 above based on the number of years in the tested grades in North Carolina.
7) If the student's performance on the current assessment is equal to or better than the minimum from the previous step, include the student in the percent proficient calculation to compare against the state's AMOs.

To illustrate, assume a student enters North Carolina in the fourth grade. The student scores below proficient in the current school year in reading. This child's known test scores are listed below.

Grade	3 EOG	4 EOG	5 EOG
Developmental Score	Not in NC	229	241
C-scale score		-2.68	-1.98

Because the student's first full year in the state is the fourth-grade year, the student will need to be on trajectory to be proficient by the end of the seventh grade and thus on the seventh-grade EOG for reading. The developmental score for seventh-grade reading equivalent to proficient is 252. The associated c-scale score is -1.00.

Because the student was not in the state for the third-grade test, the fourth-grade EOG score will be used as the baseline. The difference between the baseline and proficient on the seventh-grade test in terms of c-scale scores is 1.68 (difference between 2.68 and 1.00). For the current year (fifth grade, the second year in the state), the student must perform well enough on the test to have 33 percent less difference between the c-scale score for proficiency and his baseline (fourth-grade EOG) c-scale score (divide 1.68 by 3 = 0.56).

For this to be true, the child would need to score at least -2.12 (difference between 2.68 and 0.56). The child's actual c-scale score is -1.98 which means the child met the standard to be deemed on trajectory for the current year and thus will be included in the percent of students on trajectory or proficient for comparison to the AMO for the school as a whole and any subgroups to which the child may belong.

Tennessee

Adapted from "Evaluation of the 2005–06 Growth Model Pilot Program," U.S. Department of Education, 2009.

Tennessee uses a projection model to assess student growth. The state uses a student's history of test scores in an equation to project or predict that student's future score. To complete this process, previous cohorts of student scores are used in generating a prediction equation that can be applied to the current cohort of students. For example, last year a cohort of sixth-graders in Tennessee were tested on the state reading exam. These same sixth-graders also had scores from the state reading exam for grades 3, 4, and 5. The scores on the sixth-grade reading test are placed in a matrix called **Y**. The reading scores for grades 3–5 are placed in a matrix called **X**. All the reading scores from grades 3–6 are combined into a design matrix called **XY**. The matrices are used in a statistical procedure to generate a covariance matrix called **C** with submatrices $\mathbf{C_{XX}}$ and $\mathbf{C_{XY}}$ ($\mathbf{C_{YX}} = \mathbf{C_{XY}}^{\mathbf{T}}$) and $\mathbf{C_{YY}}$. These submatrices are used for various statistical functions but primarily in calculations for $\mathbf{b} = \mathbf{C_{XX}}^{\mathbf{-1}}\mathbf{C_{YX}}$ to generate the regression coefficients $b_1, b_2, \ldots b_N$. For example, the projected score is computed using variations of the following equation:

$$\text{Projected Score} = M_Y + b_1(X_1 - M_1) + b_2(X_2 - M_2) + \ldots + = M_Y + \mathbf{x}_i^T\mathbf{b}$$

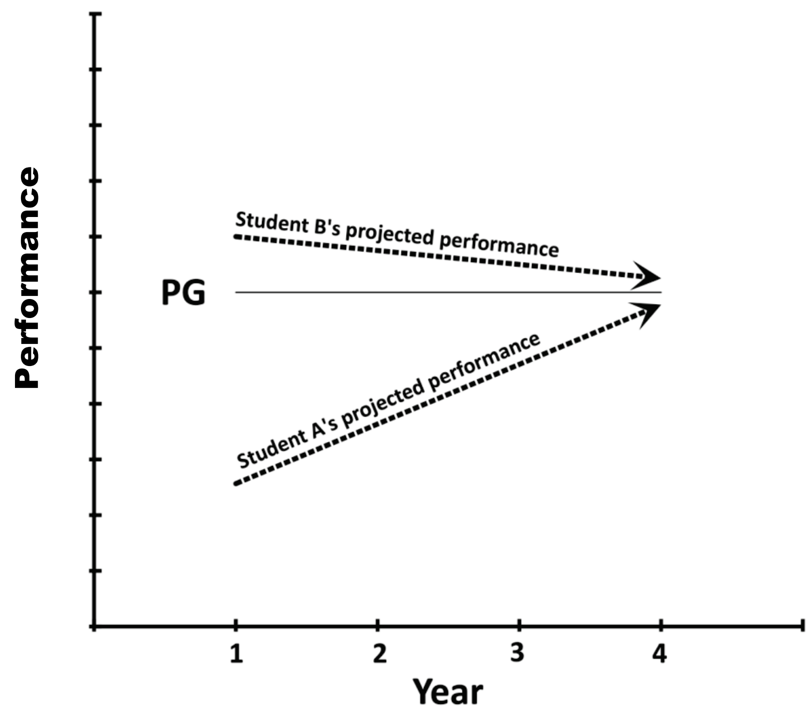
where M_Y, M_1, etc. are estimated mean scores for the "future test score" or response variable (Y). The previous test scores can also be referred to as the "predictor variables." To complete projected scores in the equation you make the following substitutions:

M_Y = estimated mean score on test
$b_1, b_2, \ldots b_N$ = regression coefficients used to predict performance
$X_1, X_2, \ldots X_N$ = previous reading scores
$M_1, M_2, \ldots M_N$ = average school reading scores

The Tennessee model includes a statewide "average schooling effect," which is obtained by calculating the mean scores for each grade of a particular school and then averaging those means over all schools within the state. It is intended to account for the fact that a current school has no control over the effectiveness of the schools that their students will attend in the future (thus potentially affecting the student's growth). The average schooling effect assumes that each student will have the "average schooling experience" of all Tennessee schools.

Tennessee's model projects scores for all students, estimating each student's performance in reading and math in three years. Each student's projection is based upon his or her available test scores. For example, student A has reading scores for 2003, 2004, and 2005 whereas student B has reading scores for 2003 and 2005. In both cases, projected scores are computed using the equation described above using student-specific values.

Exhibit B.11
Illustration of Tennessee's Method for Determining Whether a Student Is On-Track Toward Proficiency in Reading



In Exhibit B.11, student A is currently below proficient (year 1) but projected to be proficient in year 4. In Tennessee's growth model, student A would be considered proficient in year 1 for AYP determinations and in each of the succeeding years if he or she continues to be on this trajectory to proficient. This is symbolized in how student A's projection is heading towards the performance goal line (the line marked PG in Exhibit B.11).

Student B is missing data for 2004. Tennessee addresses the issue of missing test scores by using the regression coefficients from ($b = C_{XX}^{-1}C_{YX}$) and constants in the equation described above to fill in the missing test scores. Thus, all relevant student data are included in projecting score for a student. Using the available data for student B, a growth trajectory is developed that shows a projected decline in performance, though the student is projected to remain above proficient. Because student B is projected to remain above proficient by year 4, student B is proficient for the current year AYP determinations. If student B's projection indicated he or she would fall below proficient, student B would be considered non-proficient in the current year AYP determinations.

Appendix C:
Supplemental Exhibits

Exhibit C.1
Annual Measurable Objectives (AMOs) for Reading and Math, by State, 2006–07 and 2007–08 School Years

State	Reading		Math	
	2006–07	**2007–08**	**2006–07**	**2007–08**
Alaska	71.48%	77.18%	57.61%	66.09%
Arizona	44.84%	55.86%	38.44%	50.74%
Arkansas	45.49%	53.28%	41.17%	49.58%
Delaware	62.00%	68.00%	41.00%	50.00%
Florida	51.00%	58.00%	56.00%	62.00%
Iowa	67.94%	74.33%	68.10%	74.47%
North Carolina	56.05%	40.85%	68.30%	72.80%
Ohio	71.11%	76.87%	55.31%	64.26%
Tennessee	86.50%	91.00%	77.00%	84.50%

Exhibit reads: In 2006–07, schools in Alaska needed to have 71.48 percent or more of the students in each reporting group (all students plus, if present in sufficient numbers, racial or ethnic groups, LEP students, students with disabilities, and low income students) scoring at or above the proficiency cut score in order to make AYP. This AMO increased to 77.18 percent in 2007–08.

Note: A few states employed a single AMO across all grades. For the states using different AMOs for each grade, the AMO's have been averaged across grades.

Source: Alaska, Arizona, Arkansas, Delaware, Florida, Iowa, North Carolina, Ohio, and Tennessee Consolidated State Application Accountability Workbooks.

Exhibit C.2
Number of Eligible Schools and Percentage in Which Math AMO Was Met Because of Growth Model Results, by All Students and Racial/Ethnic Reporting Groups, 2007–08

Pilot States	All Students	White	Black, Non-Hispanic	Hispanic	Asian/ Pacific Islander	American Indian/Alaskan Native
	Number of Eligible Schools*					
All Nine States	1,818	371	1,924	738	47	181
Alaska	74	3	3	0	8	68
Arizona	4	37	59	45	37	99
Arkansas	49	9	77	4	0	0
Delaware**	5	NA	NA	NA	NA	NA
Florida	849	106	917	486	1	0
Iowa	71	27	41	28	1	3
North Carolina	198	2	315	104	0	11
Ohio	529	171	487	69	0	0
Tennessee	39	16	25	2	0	0
	Percent of Eligible Schools					
All Nine States	22%	46%	16%	18%	0%	0%
Alaska	0%	0%	0%	NA	0%	0%
Arizona	0%	0%	0%	0%	0%	0%
Arkansas	0%	0%	0%	0%	NA	NA
Delaware	0%	NA	NA	NA	NA	NA
Florida	9%	10%	7%	15%	0%	NA
Iowa	27%	30%	12%	11%	0%	0%
North Carolina	1%	0%	0%	<1%	NA	0%
Ohio	53%	81%	49%	77%	NA	NA
Tennessee	51%	88%	52%	0%	NA	NA

Exhibit reads: Across all nine states, among the schools in which the "all students" reporting group did not reach the mathematics AMO by either status or safe-harbor, that group did reach the AMO in 1,818 schools when the growth model criteria were applied. The number reaching the AMO by growth represented 22 percent of the eligible schools.

* "Eligible schools" means schools that did not make AYP by status or safe-harbor and had the grades required to be eligible for growth.

** ED*Facts* did not include subgroup information for Delaware in 2007–08.

NA: Not applicable due to no eligible schools.

Source: U.S. Department of Education, ED*Facts*.

Exhibit C.3
Number of Eligible Schools and Percentage in Which Math AMO Was Met Because of Growth Model Results, by Low-SES, SWD, and LEP Reporting Groups, 2007–08

Pilot States	Number of Eligible Schools*		
	Economically Disadvantaged Students	Students With Disabilities (SWD)	Limited English Proficient (LEP) Students
All Nine States	2,563	3,018	830
Alaska	81	42	53
Arizona	71	282	157
Arkansas	79	82	8
Delaware**	NA	NA	NA
Florida	1,073	1,184	460
Iowa	143	141	26
North Carolina	363	397	82
Ohio	719	877	43
Tennessee	34	13	1
	Percent of Eligible Schools		
All Nine States	24%	16%	9%
Alaska	0%	0%	0%
Arizona	0%	0%	0%
Arkansas	0%	0%	0%
Delaware	NA	NA	NA
Florida	9%	2%	11%
Iowa	16%	14%	19%
North Carolina	<1%	0%	1%
Ohio	67%	51%	40%
Tennessee	50%	15%	0%

Exhibit reads: Across all nine states, among the schools in which the “economically disadvantaged” reporting group did not reach the math AMO by either status or safe-harbor, that group did reach the AMO in 2,563 schools when the growth criteria were applied. The number reaching the AMO by growth represented 24 percent of the eligible schools.

* “Eligible schools” means schools that did not make AYP by status or safe-harbor and had the grades required to be eligible for growth.

** ED*Facts* did not include subgroup information for Delaware in 2007–08.

NA: Not applicable due to no eligible schools.

Source: U.S. Department of Education, ED*Facts*.

Exhibit C.4
Numbers of Schools at Each AYP Status by *NCLB* School Improvement Status, by State, 2007–08

Pilot States	AYP Status	Identified for Improvement or Under Corrective Action	Planning to Restructure or Restructuring	Not Identified for Improvement	Total
All Eight* States	Making AYP by Status	129	13	4,450	4,592
	Making AYP by Safe-Harbor	230	43	1,016	1,289
	Making AYP by Growth	247	27	962	1,236
	Not Making AYP	1,442	836	4,055	6,333
	Total	2,048	919	10,483	13,450
Arizona	Making AYP by Status	33	2	985	1,020
	Making AYP by Safe-Harbor	12	2	83	97
	Making AYP by Growth	3	1	4	8
	Not Making AYP	130	35	206	371
	Total	178	40	1,278	1,496
Arkansas	Making AYP by Status	22	6	103	131
	Making AYP by Safe-Harbor	54	4	311	369
	Making AYP by Growth	4	0	48	52
	Not Making AYP	133	43	162	338
	Total	213	53	624	890
Delaware	Making AYP by Status	3	0	120	123
	Making AYP by Safe-Harbor	0	0	0	0
	Making AYP by Growth	0	0	5	5
	Not Making AYP	20	15	20	55
	Total	23	15	145	183
Florida	Making AYP by Status	13	4	499	516
	Making AYP by Safe-Harbor	18	23	75	116
	Making AYP by Growth	8	13	132	153
	Not Making AYP	305	595	1,595	2,495
	Total	344	635	2,301	3,280
Iowa	Making AYP by Status	0	0	678	678
	Making AYP by Safe-Harbor	3	0	40	43
	Making AYP by Growth	5	1	17	23
	Not Making AYP	30	8	316	354
	Total	38	9	1,051	1,098
North Carolina	Making AYP by Status	10	0	378	388
	Making AYP by Safe-Harbor	55	5	275	335
	Making AYP by Growth	0	0	0	0
	Not Making AYP	309	44	1,214	1,567
	Total	374	49	1,867	2,290
Ohio	Making AYP by Status	20	0	763	783
	Making AYP by Safe-Harbor	50	4	109	163
	Making AYP by Growth	225	12	736	973
	Not Making AYP	452	89	417	958
	Total	747	105	2,025	2,877
Tennessee	Making AYP by Status	28	1	924	953
	Making AYP by Safe-Harbor	38	5	123	166
	Making AYP by Growth	2	0	20	22
	Not Making AYP	63	7	125	195
	Total	131	13	1,192	1,336

Note: Exhibit C.4 presents the counts underlying Exhibit 23 ("Numbers of Schools Making AYP by Status or Safe-Harbor, and Percentage Increase in Schools Making AYP Due to Growth, by *NCLB* School Improvement Status, 2007–08").
*Alaska did not report school improvement status to ED*Facts* for the 2007–08 school year.
Source: U.S. Department of Education, ED*Facts*

Exhibit C.5
Numbers of Schools at Each AYP Status by School Poverty Concentration, by State, 2007–08

Pilot States	AYP Status	Low Poverty	Medium Poverty	High Poverty	Total
All Nine States	Making AYP by Status	1,733	2,536	547	4,816
	Making AYP by Safe-Harbor	213	866	279	1,358
	Making AYP by Growth	352	726	168	1,246
	Not Making AYP	897	3,883	1,774	6,554
	Total	3,195	8,011	2,768	13,974
Alaska	Making AYP by Status	210	0	12	222
	Making AYP by Safe-Harbor	66	0	4	70
	Making AYP by Growth	0	0	0	0
	Not Making AYP	190	0	13	203
	Total	466	0	29	495
Arizona	Making AYP by Status	312	482	225	1,019
	Making AYP by Safe-Harbor	3	38	56	97
	Making AYP by Growth	1	3	4	8
	Not Making AYP	31	106	234	371
	Total	347	629	519	1,495
Arkansas	Making AYP by Status	4	93	34	131
	Making AYP by Safe-Harbor	21	296	52	369
	Making AYP by Growth	2	43	7	52
	Not Making AYP	3	221	114	338
	Total	30	653	207	890
Delaware	Making AYP by Status	29	90	4	123
	Making AYP by Safe-Harbor	0	0	0	0
	Making AYP by Growth	0	5	0	5
	Not Making AYP	4	47	4	55
	Total	33	142	8	183
Florida	Making AYP by Status	255	244	17	516
	Making AYP by Safe-Harbor	17	71	28	116
	Making AYP by Growth	28	93	32	153
	Not Making AYP	316	1,608	571	2,495
	Total	616	2,016	648	3,280
Iowa	Making AYP by Status	253	423	2	678
	Making AYP by Safe-Harbor	12	28	3	43
	Making AYP by Growth	7	15	1	23
	Not Making AYP	55	266	33	354
	Total	327	732	39	1,098
North Carolina	Making AYP by Status	146	221	17	384
	Making AYP by Safe-Harbor	41	250	44	335
	Making AYP by Growth	0	0	0	0
	Not Making AYP	217	1,031	311	1,559
	Total	404	1,502	372	2,278
Ohio	Making AYP by Status	409	354	35	798
	Making AYP by Safe-Harbor	49	85	29	163
	Making AYP by Growth	313	562	108	983
	Not Making AYP	80	527	377	984
	Total	851	1,528	549	2,928
Tennessee	Making AYP by Status	115	629	201	945
	Making AYP by Safe-Harbor	4	98	63	165
	Making AYP by Growth	1	5	16	22
	Not Making AYP	1	77	117	195
	Total	121	809	397	1,327

Note: Exhibit C.5 presents the counts underlying Exhibit 24 ("Numbers of Schools Making AYP by Status or Safe-Harbor, and Percentage Increase in Schools Making AYP Due to Growth, by School Poverty Concentration, 2007–08").

Source: U.S. Department of Education, ED*Facts*.

Exhibit C.6
Numbers of Schools at Each AYP Status by School Minority Concentration, by State, 2007–08

Pilot States	AYP Status	Low Minority	Medium Minority	High Minority	Total
All Nine States	Making AYP by Status	3,154	1,145	433	4,732
	Making AYP by Safe-Harbor	740	408	212	1,360
	Making AYP by Growth	895	222	113	1,230
	Not Making AYP	1,830	2,854	1,773	6,457
	Total	6,619	4,629	2,531	13,779
Alaska	Making AYP by Status	95	65	57	217
	Making AYP by Safe-Harbor	28	13	29	70
	Making AYP by Growth	0	0	0	0
	Not Making AYP	38	74	88	200
	Total	161	152	174	487
Arizona	Making AYP by Status	364	452	168	984
	Making AYP by Safe-Harbor	4	47	45	96
	Making AYP by Growth	1	3	4	8
	Not Making AYP	29	106	229	364
	Total	398	608	446	1,452
Arkansas	Making AYP by Status	75	37	10	122
	Making AYP by Safe-Harbor	268	89	12	369
	Making AYP by Growth	28	19	3	50
	Not Making AYP	110	152	69	331
	Total	481	297	94	872
Delaware	Making AYP by Status	31	86	6	123
	Making AYP by Safe-Harbor	0	0	0	0
	Making AYP by Growth	0	5	0	5
	Not Making AYP	5	41	9	55
	Total	36	132	15	183
Florida	Making AYP by Status	246	182	77	505
	Making AYP by Safe-Harbor	29	54	33	116
	Making AYP by Growth	31	78	41	150
	Not Making AYP	482	1,258	682	2,422
	Total	788	1,572	833	3,193
Iowa	Making AYP by Status	646	16	0	662
	Making AYP by Safe-Harbor	38	5	0	43
	Making AYP by Growth	20	3	0	23
	Not Making AYP	250	99	3	352
	Total	954	123	3	1,080
North Carolina	Making AYP by Status	282	87	14	383
	Making AYP by Safe-Harbor	167	142	29	338
	Making AYP by Growth	0	0	0	0
	Not Making AYP	403	821	352	1,576
	Total	852	1,050	395	2,297
Ohio	Making AYP by Status	726	41	13	780
	Making AYP by Safe-Harbor	128	18	16	162
	Making AYP by Growth	809	106	57	972
	Not Making AYP	449	238	266	953
	Total	2,112	403	352	2,867
Tennessee	Making AYP by Status	689	179	88	956
	Making AYP by Safe-Harbor	78	40	48	166
	Making AYP by Growth	6	8	8	22
	Not Making AYP	64	65	75	204
	Total	837	292	219	1,348

Note: Exhibit C.6 presents the counts underlying Exhibit 25 ("Numbers of Schools Making AYP by Status or Safe-Harbor, and Percentage Increase in Schools Making AYP Due to Growth, by School Minority Concentration, 2007–08").

Source: U.S. Department of Education, ED*Facts*.

Exhibit C.7
Numbers of Schools at Each AYP Status by School Urbanicity, by State, 2007–08

Pilot States		Urban Schools	Suburban Schools	Rural Schools	Total
All Nine States	Making AYP by Status	984	1,123	2,659	4,766
	Making AYP by Safe-Harbor	343	218	801	1,362
	Making AYP by Growth	209	492	531	1,232
	Not Making AYP	2,140	1,733	2,643	6,516
	Total	3,676	3,566	6,634	13,876
Alaska	Making AYP by Status	21	4	193	218
	Making AYP by Safe-Harbor	13	2	55	70
	Making AYP by Growth	0	0	0	0
	Not Making AYP	55	3	142	200
	Total	89	9	390	488
Arizona	Making AYP by Status	427	194	365	986
	Making AYP by Safe-Harbor	44	19	34	97
	Making AYP by Growth	5	1	2	8
	Not Making AYP	189	56	120	365
	Total	665	270	521	1,456
Arkansas	Making AYP by Status	13	5	104	122
	Making AYP by Safe-Harbor	60	30	279	369
	Making AYP by Growth	14	5	31	50
	Not Making AYP	86	25	220	331
	Total	173	65	634	872
Delaware	Making AYP by Status	13	57	53	123
	Making AYP by Safe-Harbor	0	0	0	0
	Making AYP by Growth	3	1	1	5
	Not Making AYP	16	24	15	55
	Total	32	82	69	183
Florida	Making AYP by Status	107	284	118	509
	Making AYP by Safe-Harbor	45	43	28	116
	Making AYP by Growth	35	92	24	151
	Not Making AYP	681	1,150	609	2,440
	Total	868	1,569	779	3,216
Iowa	Making AYP by Status	66	41	565	672
	Making AYP by Safe-Harbor	9	2	32	43
	Making AYP by Growth	1	2	20	23
	Not Making AYP	123	25	206	354
	Total	199	70	823	1,092
North Carolina	Making AYP by Status	55	60	281	396
	Making AYP by Safe-Harbor	64	52	222	338
	Making AYP by Growth	0	0	0	0
	Not Making AYP	449	206	952	1,607
	Total	568	318	1,455	2,341
Ohio	Making AYP by Status	71	332	381	784
	Making AYP by Safe-Harbor	34	46	83	163
	Making AYP by Growth	142	389	442	973
	Not Making AYP	421	233	305	959
	Total	668	1,000	1,211	2,879
Tennessee	Making AYP by Status	211	146	599	956
	Making AYP by Safe-Harbor	74	24	68	166
	Making AYP by Growth	9	2	11	22
	Not Making AYP	120	11	74	205
	Total	414	183	752	1,349

Note: Exhibit C.7 presents the counts underlying Exhibit 26 ("Numbers of Schools Making AYP by Status or Safe-Harbor, and Percentage Increase in Schools Making AYP Due to Growth, by School Urbanicity, 2007–08").

Source: U.S. Department of Education, ED*Facts*.

Exhibit C.8
Percentage of Schools Meeting the AMO Using Growth-Only and Making AYP by Other Means After Growth-Only Is Applied, by State, 2007–08

Pilot States		Making AYP by Status	Making AYP by Safe Harbor	Making AYP by Growth	Not Making AYP	Total
All Nine States	Not Making AYP by Growth-Only	1,415	771	202	5,220	7,608
	Making AYP by Growth-Only	3,080	548	1,042	930	5,600
	Total	4,495	1,319	1,244	6,150	13,208
Alaska	Not Making AYP by Growth-Only	12	19	0	83	114
	Making AYP by Growth-Only	206	51	0	119	376
	Total	218	70	0	202	490
Arizona	Not Making AYP by Growth-Only	517	75	7	330	929
	Making AYP by Growth-Only	440	17	1	19	477
	Total	957	92	8	349	1,406
Arkansas	Not Making AYP by Growth-Only	59	108	27	274	468
	Making AYP by Growth-Only	65	248	25	54	392
	Total	124	356	52	328	860
Delaware	Not Making AYP by Growth-Only	47	0	4	50	101
	Making AYP by Growth-Only	75	0	1	4	80
	Total	122	0	5	54	181
Florida	Not Making AYP by Growth-Only	211	91	111	2,419	2,832
	Making AYP by Growth-Only	305	25	42	75	447
	Total	516	116	153	2,494	3,279
Iowa	Not Making AYP by Growth-Only	279	30	15	299	623
	Making AYP by Growth-Only	363	13	8	52	436
	Total	642	43	23	351	1,059
North Carolina	Not Making AYP by Growth-Only	110	279	0	1,215	1,604
	Making AYP by Growth-Only	145	42	0	32	219
	Total	255	321	0	1,247	1,823
Ohio	Not Making AYP by Growth-Only	10	12	23	364	409
	Making AYP by Growth-Only	690	143	958	563	2,354
	Total	700	155	981	927	2,763
Tennessee	Not Making AYP by Growth-Only	170	157	15	186	528
	Making AYP by Growth-Only	791	9	7	12	819
	Total	961	166	22	198	1,347

Note: Exhibit C.8 presents the counts underlying Exhibits 31 and 32 ("Percentage of Schools Meeting AMO Using Only the Growth Model On-Track Indicator, by Standard ED*Facts* AYP Classification and State, 2007–08" and "Percentage of Schools Meeting the AMO Using Growth-Only and Making AYP by Other Means After Growth-Only is Applied, by State, 2007–08," respectively).

Source: U.S. Department of Education, ED*Facts* and the Alaska, Arizona, Arkansas, Delaware, Florida, Iowa, North Carolina, Ohio, and Tennessee state departments of education.

Appendix D:
Derivation of the Generic Projection Model Rule for Identifying On-Track Students

The slope of the projection model line in Exhibit 45 and subsequent exhibits is initially surprising in its stark contrast with alternative models. This appendix explains the derivation of the line in these figures.

The prediction equations used by the projection model are shown below. Note that they are simplified due to the standardization of the North Carolina data in the construction of a "generic" dataset.

$$\hat{R}_{g+N} = \hat{\beta}^{*}_{R1} R_{g-1} + \hat{\beta}^{*}_{R2} R_g$$

$$\hat{M}_{g+N} = \hat{\beta}^{*}_{M1} M_{g-1} + \hat{\beta}^{*}_{M2} M_g$$

The projection model line is easily calculated from these prediction equations. For example, the predicted reading score in some future grade $g + N$ is $\hat{R}_{g+N}$. This is compared with the proficient cut score at that grade, c_{g+N}. Thus, the projection model decision rule for Reading is defined by the following:

If $\hat{\beta}^{*}_{R1} R_{g-1} + \hat{\beta}^{*}_{R2} R_g \geq c_{g+N}$, then the student is on track.

The framework displays this decision rule on a plot of R_g on R_{g-1}. Thus, solving for the vertical axis, R_g, the decision line can be written in slope-intercept form as follows:

$$R_g = \left(-\frac{\hat{\beta}^{*}_{R1}}{\hat{\beta}^{*}_{R2}}\right) R_{g-1} + \left(\frac{c_{g+N}}{\hat{\beta}^{*}_{R2}}\right)$$

The contrast between the projection model line and the positive-sloped trajectory and transition matrix model decision lines is revealed by the simple fact that regression coefficients in this framework are positive, leading to a projection model line that is negative. Exhibit D.1 overviews the regression coefficients and corresponding slopes for the North Carolina data. The regression line in the framework is plotted from the Math dataset that uses prediction grades 4 and 5 to project to grade 8. As Exhibit D.1 shows, the use of an alternative grade would not change the fundamental contrast between projection and trajectory models, and in most cases the contrast would be increased.

Exhibit D.1
Regression Coefficients and Framework Slopes for North Carolina Data

Subject	Prediction Grades ($g-1$ and g)	Projection Grade	$\hat{\beta}^*_{j1}$	$\hat{\beta}^*_{j2}$	Slope $\left(-\frac{\hat{\beta}^*_{R1}}{\hat{\beta}^*_{R2}}\right)$
Reading	2 and 3	6	0.286	0.543	-0.527
Reading	3 and 4	7	0.388	0.491	-0.688
Reading	4 and 5	8	0.378	0.469	-0.806
Reading	5 and 6	8	0.383	0.487	-0.787
Reading	6 and 7	8	0.407	0.481	-0.846
Math	3 and 4	7	0.366	0.501	-0.730
Math	4 and 5	8	0.336	0.512	-0.657
Math	5 and 6	8	0.318	0.562	-0.566
Math	6 and 7	8	0.364	0.539	-0.675

The Department of Education's mission is to promote student achievement and preparation for global competitiveness by fostering educational excellence and ensuring equal access.

www.ed.gov